

Detection and characterization of pathogenic and non-pathogenic *Brachyspira*

Submitted to the College of Graduate Studies and Postdoctoral Studies of the University of Saskatchewan in partial fulfillment of the requirements for the degree of Master of Science in the Department of Veterinary Microbiology at the University of Saskatchewan.

By

LISA ANNE JOHNSON

© Copyright Lisa A. Johnson, October, 2017. All Rights Reserved.

PERMISSION TO USE

In agreement with the outlines set out by the College of Graduate Studies and Postdoctoral Studies at the University of Saskatchewan, I allow the University of Saskatchewan Libraries to make this thesis available to all interested parties. Also in accordance with the College of Graduate Studies and Research, I allow this thesis to be copied “in any manner, in whole or in part, for scholarly purposes”. This thesis may not, however, be reproduced or used in any manner for financial gain without my written consent. Any scholarly use of this thesis, in part or in whole, must acknowledge both myself and the University of Saskatchewan.

Any requests for copying or using this thesis, in any form or capacity, should be made to:

Head of Department of Veterinary Microbiology

University of Saskatchewan

Saskatoon, Saskatchewan

S7N 5B4

ABSTRACT

Brachyspira is associated with mucohaemorrhagic diarrhea in pigs, though the pathogenesis and virulence factors behind the disease are not currently understood. This thesis aimed to identify putative virulence factors of *Brachyspira* through a comparative genomics approach, and to determine the frequency of the occurrence of mixed *Brachyspira* species infections within clinical case material.

“*B. hampsonii*” strain 30446 was originally isolated from colon contents of a pig with mucohaemorrhagic diarrhea, and through inoculation trials has been demonstrated to be pathogenic in swine. “*B. hampsonii*” strain KL-180 was isolated from a Lesser Snow Goose, and is remarkably genetically similar to strain 30446. However, inoculation trials indicate “*B. hampsonii*” strain KL-180 is not pathogenic in swine. Their differing pathogenicity, despite their genetic similarity, allows comparison to identify gene content differences, which may correspond to putative virulence factors. Genome sequences for each strain were available, and were annotated. The basic genome features of strain 30446 and strain KL-180 were similar, with an average nucleotide identity (ANI) of 97%. Comparison of the annotated genomes revealed genes unique to each strain. Any gene differences could be important to disease outcome, including hypothetical proteins. “*B. hampsonii*” strain 30446, contained a set of genes absent in strain KL-180 which correspond to the Streptolysin S (*sag*) operon. The presence of this operon was confirmed with PCR and primers designed for each gene. The presence of this operon in pathogenic *Brachyspira* species, and absence in the non-pathogenic “*B. hampsonii*” strain KL-180 supports the possibility of the operon being a virulence factor and contributing to disease outcome.

Current diagnostic methods implemented on clinical case material often indicate the presence of more than one *Brachyspira* species. It is not currently known if co-infection with multiple species is common, or if the presence of multiple species contributes to *Brachyspira* pathogenesis. To investigate the frequency of the occurrence of multiple species in *Brachyspira* clinical cases, deep sequencing of the *nox* gene target was used on clinical case material from Western Canada, Mexico and Brazil. Synthetically created communities were also sequenced, to provide insight as to what levels of *Brachyspira* can be detected through this method. Pigs with clinical disease were very frequently colonized by multiple species of *Brachyspira*. “*B. hampsonii*” was detected in pigs from both Mexico and Brazil, though often at very low levels. Canadian clinical cases were more diverse than those from Mexico and Brazil, and were dominated by a range of *Brachyspira* species. Sequencing of the synthetic communities revealed a negative primer bias towards “*B. hampsonii*” strain 30599 (Clade I). Strain 30599 may therefore be present at higher levels and be more widely distributed than demonstrated by current diagnostic methods based on PCR of the *nox* gene.

The results of this thesis demonstrate that the *sag*-like operon identified through comparative genomics and found in pathogenic strains of *Brachyspira* may be a virulence factor. Additionally, mixed *Brachyspira* infections are commonplace within clinical case material, and multiple species may be involved in disease outcome.

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my supervisor Dr. Janet Hill for her patience and guidance. I cannot imagine a better supervisor, and her wealth of knowledge never ceases to amaze me. Writing #Lisasthesis was easier knowing I was lucky enough to have such great guidance. I have grown and learned so much in my time with the Hill Lab, and will always be thankful for my start as summer student in 2013!

Thanks to the rest of the Hill Lab (past and present)! Champika, thank you for your patience and kindness, and support in the lab. It was always reassuring to have you as such a helpful resource. More thanks to Bonnie, Teenus, Isha, John S., Matheus, Aline, Niradha, Jason, Mo and Salahuddin for all your camaraderie and for being wonderful people to work with! Also, special thanks to Sonu, Arinjay, and Carolina for being great officemates and always being fun to chat with.

Thanks to Dr. Aaron White for his help on my advisory committee, and Dr. Vikram Misra for his work as graduate chair.

Thanks to the rest of the Veterinary Microbiology department for your friendship, and making this a fantastic department to work in.

Last but not least, I would like to thank my family. The support, understanding and encouragement from my parents means the world to me, and I would not be where I am today without it. Special thanks to David and Sarah for being the best siblings one could ask for, and for being supportive throughout my studies

TABLE OF CONTENTS

PERMISSION TO USE	i
ABSTRACT	ii
ACKNOWLEDGEMENTS.....	iv
LIST OF TABLES	vii
LIST OF FIGURES	viii
1 INTRODUCTION AND LITERATURE REVIEW	1
1.1 BRACHYSPIRA-ASSOCIATED MUCOHAEMORRHAGIC DIARRHEA	1
1.1.1 RE-EMERGENCE OF DISEASE	2
1.1.2 SWINE ASSOCIATED BRACHYSPIRA SPP. AND THEIR PATHOGENICITY	2
1.1.3 GLOBAL DISTRIBUTION OF “B. HAMPSONII”	6
1.1.4 DIAGNOSTIC METHODS	7
1.1.5 PATHOGENESIS AND PUTATIVE VIRULENCE FACTORS	9
1.1.6 MULTIPLE BRACHYSPIRA SPP. INFECTIONS.....	11
1.2 COMPARATIVE GENOMICS.....	12
1.2.1 GENOME ANNOTATION	12
1.2.2 APPLICATIONS IN MICROBIOLOGY	17
2 OBJECTIVES	19
3 COMPARATIVE GENOMICS OF PATHOGENIC AND NON-PATHOGENIC STRAINS OF “BRACHYSPIRA HAMPSONII” AND RELATIONSHIP OF GENE CONTENT DIFFERENCES TO PUTATIVE VIRULENCE FACTORS	20
3.1 ABSTRACT.....	21
3.2 BACKGROUND.....	22
3.3 METHODS.....	23
3.3.1 BRACHYSPIRA GENOME SEQUENCES	23
3.3.2 GENOME ANNOTATION	24
3.3.3 BRACHYSPIRA ISOLATES AND CULTURE	26
3.3.4 DNA EXTRACTION AND PCR	26
3.4 RESULTS.....	29
3.4.1 GENOME ANNOTATION	29
3.4.2 CLUSTERS OF ORTHOLOGOUS GROUPS (COGs)	34
3.4.3 FUNCTIONAL COMPARISON OF “B. HAMPSONII” STRAINS 30446 AND KL-180 IN RAST	36
3.4.4 SAG-LIKE GENES IN BRACHYSPIRA	41
3.4.5 LOCATING SAGA IN BRACHYSPIRA GENOMES	43
3.4.6 SAG-LIKE GENES IN “B. HAMPSONII” AND OTHER BRACHYSPIRA SPP.	45
3.5 DISCUSSION.....	48
3.6 CONCLUSIONS.....	53
3.7 TRANSITION STATEMENT	54

4	CHARACTERIZATION OF BRACHYSPIRA COMMUNITIES FROM CLINICAL CASES OF SWINE MUCOHAEMORRHAGIC DIARRHEA	55
4.1	ABSTRACT.....	56
4.2	BACKGROUND.....	57
4.3	METHODS.....	58
4.3.1	BRACHYSPIRA ISOLATES, CULTURE AND DNA EXTRACTION	58
4.3.2	CLINICAL SAMPLES.....	59
4.3.3	CONSTRUCTION OF SYNTHETIC BRACHYSPIRA COMMUNITIES AND QUANTITATIVE REAL-TIME PCR.....	61
4.3.4	DEEP SEQUENCING AND ANALYSIS	64
4.4	RESULTS.....	65
4.4.1	NOX AMPLICON SEQUENCE SIMILARITY	65
4.4.2	LIMIT OF DETECTION OF CONVENTIONAL NOX PCR AND SYNTHETIC COMMUNITIES ..	68
4.4.3	DEEP SEQUENCING OF SYNTHETIC COMMUNITIES AND CLINICAL SAMPLES.....	71
4.4.4	BRACHYSPIRA COMMUNITIES FROM SYNTHETIC COMMUNITIES.....	72
4.4.5	BRACHYSPIRA COMMUNITIES FROM CLINICAL CASES	75
4.5	DISCUSSION.....	83
4.6	CONCLUSIONS.....	87
5	GENERAL DISCUSSION.....	89
5.1	SUMMARY AND LIMITATIONS OF THESE WORKS.....	89
5.1.1	COMPARATIVE GENOMICS OF “B. HAMPSONII” STRAINS REVEALS GENE CONTENT DIFFERENCES.....	89
5.1.2	DEEP SEQUENCING OF FECAL DNA FROM CLINICAL CASES OF BRACHYSPIRA- ASSOCIATED DISEASE REVEAL MULTIPLE BRACHYSPIRA SPP. INFECTIONS ARE COMMON	91
5.2	DISCUSSION OF FUTURE PROSPECTS	91
6	REFERENCES	93

LIST OF TABLES

Table 1.1. Summary of recognized <i>Brachyspira</i> species, and provisional species “ <i>B. hampsonii</i> ”.	5
Table 3.1. Characteristics of genome sequences used in this study.	25
Table 3.2. Primer sequences used in this study.	28
Table 3.3. Genome annotation characteristics obtained from (A) the RAST server, and (B) the JGI MGAP.	31
Table 3.4. Average Nucleotide Identity (ANI _m) values for pairwise comparisons between four <i>Brachyspira</i> spp..	33
Table 3.5. Genes associated with a subsystem in “ <i>B. hampsonii</i> ” strain 30446, and absent in “ <i>B. hampsonii</i> ” strain KL-180, based on the metabolic construction in RAST.	38
Table 3.6. Genes associated with a subsystem in “ <i>B. hampsonii</i> ” strain KL-180, and absent in “ <i>B. hampsonii</i> ” strain 30446, based on the metabolic construction in RAST.	40
Table 3.7. Genes present in <i>B. hyodysenteriae</i> strain WA1, “ <i>B. hampsonii</i> ” strain 30446 and strain 30599, but absent in “ <i>B. hampsonii</i> ” strain KL-180, as determined from annotation on JGI MGAP.	42
Table 3.8. Detection of <i>sag</i> -like genes in <i>Brachyspira</i> spp. by conventional PCR.	47
Table 4.1. Primer sequences used in this study.	63
Table 4.2. The amount of each <i>Brachyspira</i> species in the synthetic <i>Brachyspira</i> communities as determined by qPCR.	70

LIST OF FIGURES

Figure 3.1. Pangenome analysis obtained from the GView server.....	32
Figure 3.2. Distribution of COGs in <i>Brachyspira</i> genomes.....	35
Figure 3.3. sag-like operon in <i>Brachyspira</i>	44
Figure 4.1. The percent identity of 70 pairwise comparisons of full-length nox gene sequences, the first 200 bp, and the last 200 bp.....	67
Figure 4.2. <i>B. hyodysenteriae</i> , <i>B. pilosicoli</i> , “ <i>B. hampsonii</i> ” 30446, and “ <i>B. hampsonii</i> ” 30599 genomic DNA dilution curves from 10^7 to 10^0 copies/ μ l visualized on an agarose gel.....	69
Figure 4.3. Proportion of four different <i>Brachyspira</i> species present in synthetically created communities.....	73
Figure 4.4. Proportion of each community that is comprised of different <i>Brachyspira</i> species for each Mexican clinical case (n=31).....	76
Figure 4.5. Proportion of each community that is comprised of different <i>Brachyspira</i> species for each Brazilian case (n=11).....	77
Figure 4.6. Proportion of each community that is comprised of different <i>Brachyspira</i> species for each Canadian clinical case (n=47).....	79
Figure 4.7. Cluster analysis of <i>Brachyspira</i> communities based on the proportion of different <i>Brachyspira</i> species present in each sample.	81
Figure 4.8. Principal Coordinates Analysis based on Bray-Curtis dissimilarity values.	82

1 Introduction and Literature Review

1.1 *Brachyspira*-associated mucohaemorrhagic diarrhea

Swine dysentery, or mucohaemorrhagic diarrhea, was first described in 1921 (Whiting et al., 1921), though the causative agent was not identified until 1972. The agent was identified as a spirochete and was named *Treponema hyodysenteriae* (Glock et al., 1974; Harris et al., 1972). Classification of the genus of this spirochete was changed to *Serpula*, *Serpulina*, and finally to its current classification, *Brachyspira* (Ochiai et al., 1997; Stanton, 1992; Stanton et al., 1991). *B. hyodysenteriae* is considered the hallmark causative agent behind swine dysentery, although there are other species of *Brachyspira* with varying levels of pathogenicity in pigs.

B. hyodysenteriae is a spirochete that is Gram-negative, oxygen tolerant, with 14-18 periplasmic flagella per cell (Ochiai et al., 1997). Infection by *B. hyodysenteriae* results in swine dysentery, and is characterized by mucohaemorrhagic diarrhea, colitis, and lesions within the large intestine. Transmission of swine dysentery occurs via the ingestion of feces from an infected animal, and it is difficult to eradicate from swine production systems once *Brachyspira* has been introduced. Herds that are infected can display morbidity of 90%, and mortality of up to 30% (Harris and Glock, 1971). Infected herds are not economically desirable as swine dysentery results in poor feed conversion and growth rates, increasing size variation of pigs within the herd (Harris and Glock, 1971).

1.1.1 Re-emergence of disease

Due to increased biosecurity and the use of antibiotics, swine dysentery became rare in Canada but re-emerged in the late 2000s. Swine dysentery-like disease was described on a farm in Western Canada in 2009, but *B. hyodysenteriae* could not be detected in clinical samples (Harding et al., 2010). Using the NADH-oxidase (*nox*) gene target, sequencing of partial *nox* gene PCR amplicons identified the causative agent of this re-emergence to be an atypical *Brachyspira* sp. only 92% identical in *nox* sequence to any known species (Harding et al., 2010). Further analysis of clinical cases demonstrated an association between detection of this atypical *nox* sequence with mucohaemorrhagic diarrhea in pigs free from classically associated pathogens, such as *B. hyodysenteriae* (Harding et al., 2012).

At the same time, atypical *Brachyspira* sp. associated with the re-emergence of disease were also described in the United States, and were related to the same atypical *Brachyspira* observed in Canada. Upon further investigation of the novel species through sequencing methods of the *nox* and rRNA genes, this new species could be resolved and the name “*B. hampsonii*” has been proposed (Chander et al., 2012). “*B. hampsonii*” can be further divided into clades I and II based on sequence data, however, clinical disease associated with either clade is indistinguishable (Costa et al., 2014b; Rubin et al., 2013a).

1.1.2 Swine associated *Brachyspira* spp. and their pathogenicity

Along with *B. hyodysenteriae*, there are six additional formally recognized *Brachyspira* species and species that lack formal nomenclature, that all vary in their pathogenicity to pigs (Table

1.1). *B. hyodysenteriae* is the traditionally associated causative agent of swine dysentery, and is associated with clinically severe disease (Harris et al., 1972). *B. pilosicoli* is pathogenic, though causes a mild colitis and diarrhea. Infection causes porcine intestinal spirochaetosis, characterized by watery mucoid diarrhea (Trott et al., 1996). *B. pilosicoli* also has been isolated and associated with similar disease outcomes of watery feces in poultry and humans (Margawani et al., 2004; McLaren et al., 1997; Stephens and Hampson, 2002). Disease in different animal species could be significant epidemiologically, and is an important zoonotic consideration. *B. murdochii* was considered a commensal in swine, though it has been associated with a mild colitis and watery diarrhea (Jensen et al., 2010; Osorio et al., 2013; Weissenböck et al., 2005). *B. murdochii* can account for a high prevalence of *Brachyspira* species detected in pigs (Komarek et al., 2009; Osorio et al., 2013), and it has been observed in the gut of healthy pigs (Patterson et al., 2013). The significance of *B. intermedia*, or the degree of disease it causes, is unclear. Originally *B. intermedia* was thought to be non-pathogenic in pigs, though has been associated with disease. The significance of *B. intermedia* is further complicated by great levels of diversity within the species which leads researchers to believe isolates currently defined as *B. intermedia* may contain multiple distinct species (Phillips et al., 2010).

“*B. hampsonii*” is a novel species associated with the re-emergence of mucohaemorrhagic diarrhea in swine in Canada and the United States. Koch’s postulates were fulfilled by swine inoculation trials where pigs inoculated with clinical isolates developed mucohaemorrhagic diarrhea, and “*B. hampsonii*” was isolated from infected animals (Burrough et al., 2012b). “*B. hampsonii*” strains isolated from pigs in the US were inoculated as pure culture into mice and pigs and produced lesions similar to *B. hyodysenteriae*, indicating pathogenicity of the atypical species

(Burrough et al., 2012a; Burrough et al., 2012b). Further experimental inoculation trials established the pathogenicity of “*B. hampsonii*” clade I (Costa et al., 2014b), and “*B. hampsonii*” clade II (Rubin et al., 2013a; Wilberts et al., 2014). Clinical symptoms following the infection of “*B. hampsonii*” clade I and clade II were indistinguishable from those caused by *B. hyodysenteriae* in swine dysentery.

Table 1.1. Summary of seven recognized *Brachyspira* species, and provisional species “*B. hampsonii*”.

Taxon	Host	Pathogenicity to pigs
<i>B. hyodysenteriae</i>	Pig	Pathogenic
<i>B. innocens</i>	Pig	Non-pathogenic
<i>B. intermedia</i>	Pig	Non-pathogenic
<i>B. pilosicoli</i>	Pig, poultry, human	Pathogenic
<i>B. murdochii</i>	Pig	Non-pathogenic
<i>B. aalborgi</i>	Human	Non-pathogenic
<i>B. alvinipulli</i>	Poultry	Non-pathogenic
“ <i>B. hampsonii</i> ”	Pig	Pathogenic

1.1.3 Global distribution of “*B. hampsonii*”

Brachyspira survives in different environments and hosts, complicating the understanding of the distribution of these bacteria. Knowledge of the distribution of *Brachyspira*, including “*B. hampsonii*”, in Canada is incomplete as it generally relies on passive surveillance of diagnostic samples submitted by producers and veterinarians. Reservoirs for *Brachyspira* can include, but are not limited to, pigs, water, soil, rodents, and migrating waterfowl (Backhans et al., 2010; Boye et al., 2001; Jansson et al., 2004; Joens and Kinyon, 1982; Martinez-Lobo et al., 2013; Oxberry et al., 1998; Phillips et al., 2009; Rasback et al., 2007; Rubin et al., 2013b). Exacerbating the problem of *Brachyspira* control is that *Brachyspira* can persist in the environment outside of their host, and can survive in feces for 48 days at temperatures between 0 and 10°C (Chia and Taylor, 1978).

“*B. hampsonii*” is primarily accountable for the re-emergence of disease in North America, and has been detected in pigs from both Canada and the United States (Chander et al., 2012; Harding et al., 2010; Mirajkar et al., 2015; Rubin et al., 2013a). “*B. hampsonii*” has also been isolated from pigs in Belgium and Germany (Mahu et al., 2014; Rohde et al., 2014). At the time of writing, the distribution of “*B. hampsonii*” in swine producing areas of South America and Asia is not known.

Pig-associated *Brachyspira* spp. have been isolated from migrating waterfowl in Europe (Aller-Moran et al., 2016; Martinez-Lobo et al., 2013). One strain isolated from migratory waterfowl, AIS50, was used to experimentally inoculate pigs. All inoculated pigs shed “*B. hampsonii*”, and there was transmission of “*B. hampsonii*” from inoculated pigs to sentinel pigs, although clinical signs of swine dysentery occurred in only one of the five inoculated pigs (Aller-

Moran et al., 2016). In the Canadian Arctic, “*B. hampsonii*” was isolated from Lesser Snow Geese (Rubin et al., 2013a). One strain isolated from a goose in the Arctic, KL-180, was used to inoculate pigs, and results indicate the bacteria colonized the gut and was shed in feces, but did not produce clinical signs in pigs (Rubin et al., 2013a). Remarkably, “*B. hampsonii*” strain KL-180 and the pathogenic “*B. hampsonii*” strain 30446 are genetically similar despite their differences in pathogenicity. Based on *nox* and *cpn60* gene sequences, “*B. hampsonii*” strain KL-180 and strain 30446 were 99.9% and 99.5% similar, respectively (Rubin et al., 2013a). Migratory waterfowl may play a role in the distribution of *Brachyspira* species, and it is possible *Brachyspira* species that have colonized birds have adapted to their environment, resulting in changes to the pathogenic potential of these bacteria in pigs.

1.1.4 Diagnostic methods

Identification of the species present during *Brachyspira*-related disease is important in furthering our understanding of the disease and developing methods to prevent or treat the disease. Diagnostics can use a combination of clinical signs, gross and histopathology, culture of spirochaetes, biochemical assays, and PCR methods (Burrough, 2016). Methods such as microscopy, haemolysis strength, and biochemical tests are useful to identify *Brachyspira* as present, but do not provide high resolution for differentiating *Brachyspira* species (Chander et al., 2012; Mahu et al., 2016; Perez et al., 2016). Diagnostics often use PCR which provides species level differentiation capabilities (Burrough, 2016).

Matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) systems are utilized for the identification of bacteria (Bizzini and Greub, 2010), and are

becoming more available in diagnostic laboratories. MALDI-TOF MS has the advantage of identifying bacterial isolates accurately and inexpensively, though initial set-up costs can be limiting (Bizzini and Greub, 2010). It has been demonstrated to be a useful method for identification of *Brachyspira* species, however it is not sufficient for differentiating between “*B. hampsonii*” clades (Calderaro et al., 2013; Warneke et al., 2014).

PCR is important for the detection and identification of *Brachyspira* from clinical cases, given it is a quick method to provide results. The NADH-oxidase (*nox*) gene is a commonly used genetic marker for *Brachyspira*, as it provides greater differentiation of *Brachyspira* species than other universal gene targets, such as the 16S rRNA gene (Chander et al., 2012). Genus-specific PCR primers amplify the *nox* sequence from *Brachyspira* spp., allowing for species level identification from sequencing the amplicon (Rohde et al., 2002). *nox* PCR can be performed directly on a sample, or in combination with selective culture methods. PCR protocols are also available for the detection of individual species, such as *B. hyodysenteriae* and *B. pilosicoli* (La et al., 2006; La et al., 2003). Quantitative PCR (qPCR) assays designed to be species-specific are used in *Brachyspira* diagnostics (Rubin et al., 2013a; Song and Hampson, 2009).

Assays such as those described above are useful to determine the presence of a specific species, but do not provide insight into other *Brachyspira* that may be present. A combination of the above methods are utilized in *Brachyspira* diagnostics, often including selective culture followed by PCR and sequencing, genus-specific PCR followed by sequencing, and species-specific qPCR. Using multiple methods can result in the identification of multiple *Brachyspira* species from one case, despite one species often been attributed to causing disease. Due to the

nature of multiple *Brachyspira* being identified from single clinical cases, there is need for diagnostic methods that simultaneously detect multiple species.

1.1.5 Pathogenesis and putative virulence factors

Brachyspira-associated disease, or swine dysentery, occurs by the bacteria colonizing the large intestine where they are found on the luminal surface and within the crypts of the intestine. The pathogenesis of swine dysentery is not fully understood, though virulence factors have been identified that may contribute to the disease outcome, including haemolysins, lipoligosaccharides, motility of the bacteria, and chemotaxis to mucus and resistance to oxygen toxicity (Kennedy and Yancey, 1996; Milner and Sellwood, 1994; Muir et al., 1992; Naresh and Hampson, 2010; Nibbelink and Wannemuehler, 1991; Nuessen et al., 1983; Rosey et al., 1996; Stanton et al., 1999).

Brachyspira have periplasmic flagella that aid motility and allow the bacteria to move through mucus. They have a snake-like motion which aids movement through highly viscous materials and flagellar function is important in allowing penetration of mucus. Movement through the mucus layer is vital in allowing for the bacteria to access the epithelial cells within the crypts of the large intestine. Movement of *Brachyspira* is also influenced by its strong chemotaxis to porcine mucins (Kennedy and Yancey, 1996; Milner and Sellwood, 1994; Naresh and Hampson, 2010). Resistance to oxygen toxicity is key to the survival of anaerobic *Brachyspira*, and movement into the crypts and under the mucus protects from exposure to oxygen. Another important resistance mechanism to oxygen toxicity is NADH-oxidase (*nox*), which consumes oxygen. It has been demonstrated that *nox* mutants in *B. hyodysenteriae* cause a less severe disease in pigs (Stanton et al., 1999).

Haemolysis is the breakdown of blood cells, which can be observed on agar plates containing blood as zones of clearing. It has been suggested that strong β -haemolysis is demonstrative of pathogenic *Brachyspira*, while weak β -haemolysis is a characteristic of low pathogenicity (Burrough et al., 2012b; Hyatt et al., 1994). Generally, species that are pathogenic in swine and cause severe clinical symptoms, such as *B. hyodysenteriae* and “*B. hampsonii*”, are defined as being strongly haemolytic. Currently, there are seven putative haemolysin genes that have been identified in *B. hyodysenteriae*. *tlyA*, *tlyB* and *tlyC* genes encode haemolysins that are present in *B. hyodysenteriae* (Muir et al., 1992; ter Huurne et al., 1994). In a previous study, *tlyA* mutants that did not have the functioning hemolysin gene did not result in clinical symptoms when inoculated into pigs, and pigs that were previously colonized by the *tlyA* mutant strain had partial protection when challenged with virulent *B. hyodysenteriae* (Hyatt et al., 1994). Another haemolysin gene that has been identified is *hlyA*, which is also found in *B. hyodysenteriae* but is distinct from *tlyA* to *tlyC*. Haemolytic activity of *hlyA* was demonstrated by using haemolysis negative *E. coli* which gained β -haemolytic activity with the presence of the gene (Hsu et al., 2001). *B. hyodysenteriae* strain WA1 was the first representative of the genus *Brachyspira* to have its complete genome sequence determined, and predicted proteins from the genome elucidated potential virulence factors including haemolysins (Bellgard et al., 2009). *tlyA*, *tlyB*, *tlyC* and *hlyA* were located within the genome, as well as two genes involved in encoding putative haemolysin III and one gene encoding a putative haemolysin CBS domain containing protein (Bellgard et al., 2009).

Previous studies suggest that in addition to specific virulence factors, the resident microbiota and *Brachyspira* load may play roles in disease outcome. Experiments in which

gnotobiotic pigs were inoculated with *B. hyodysenteriae* alone did not produce lesions consistent with swine dysentery (Harris et al., 1978). Another study demonstrated again that *B. hyodysenteriae* alone was not sufficient to cause disease in gnotobiotic pigs, though inoculation of *B. hyodysenteriae* with various anaerobic bacteria did produce clinical signs (Whipp et al., 1979). *Brachyspira* are primarily responsible for swine dysentery, though one or more other anaerobic bacteria may be responsible in determining susceptibility of pigs to disease (Costa et al., 2014a).

1.1.6 Multiple *Brachyspira* spp. infections

While the microbiota may play a role in disease outcome, it has also become evident in clinical cases that multiple *Brachyspira* species are often present. Routine diagnostics often employ multiple methods to detect and/or identify *Brachyspira*, which often are a combination of culture, and PCR methods. Different methods will often identify different *Brachyspira* species, indicating the presence of multiple species. A study describing the prevalence of weakly haemolytic *Brachyspira* in pigs found that concurrent infections with two or more *Brachyspira* spp. was common, and weakly haemolytic species may contribute to colonic pathology (Komarek et al., 2009). Individual fecal samples of chickens were also observed to contain multiple *Brachyspira* species (Jansson et al., 2008).

The traditionally causative and pathogenic agent, *B. hyodysenteriae*, has also been isolated from pigs without clinical disease (La et al., 2016a) demonstrating that *B. hyodysenteriae* alone may not always be able to cause disease. In an experimental infection model, inoculation with “*B. hampsonii*” clade II strain 30446, which is also considered to be pathogenic, caused 8/12 inoculated pigs to develop mucohaemorrhagic diarrhea (Rubin et al., 2013a). The remaining 4 pigs

did not develop clinical signs of swine dysentery. While “*B. hampsonii*” strain 30446 may be pathogenic it may require additional organisms to cause disease and mixed *Brachyspria* species infections may be important. It is not known whether mixed infections are of clinical significance in regards to swine dysentery, and whether multiple species may contribute to the pathogenesis of the disease.

1.2 Comparative genomics

High throughput “next generation” sequencing is becoming an economically viable, and informative tool to investigate bacterial genomes (Loman et al., 2012). With vast amounts of sequence data becoming available, comparative analysis of sequence data becomes possible. Comparative genomics uses computer analysis to compare the genome sequences of different organisms, identifying similarities or differences which can provide insight into bacterial physiology, evolution, and pathogenicity. In order for functional comparison of genome information, an important step is identifying the features of the genome through genome annotation.

1.2.1 Genome annotation

Genome annotation aims to identify the genes encoded within the DNA sequence of an organism and their function. Genome annotation can be performed at the nucleotide-level, protein-level, and process-level (Stein, 2001). Identifying the locations of genes, genetic markers, and important genetic landmarks is the first step in the nucleotide sequence analysis. Gene finding occurs by identifying open reading frames (ORFs), and often employs a computer program that

performs a six-frame translation and locates all potential ORFs (Stein, 2001). Once the location of putative genes has been determined, the next step is to identify putative functions of the encoded proteins. The difficulty with identifying function is that during evolution, copies of the same gene can diverge creating a family of related proteins, known as paralogues (Stein, 2001). One system to identify and cluster groups of orthologous proteins is the COG system (Clusters of Orthologous Groups) (Tatusov et al., 2001), though many proteins cannot be classified through this automated process due to a lack of information. In a study of a range of prokaryotic genomes aimed at elucidating how many proteins could be classified into COG categories, 60 to 86% of genes per genome were classified into a COG category, leaving 14 to 40% of the genes unclassified (Galperin et al., 2015). The last step of genome annotation is relating the genes and their functions to biological process, which can involve techniques beyond computational work (Stein, 2001). Laboratory experiments confirming gene expression and function may be necessary. Many genome annotation servers and programs exist, but two are used in this thesis and are described below: Rapid-Annotation using Subsystems Technology (RAST) and the Joint Genome Institute's (JGI) Microbial Genome Annotation Pipeline (MGAP).

1.2.1.1 *RAST*

Rapid Annotation using Subsystems Technology (RAST) is an automated annotation service for archaeal and bacterial genomes provided by the National Microbial Pathogen Data Resource (NMPDR) (<http://rast.nmpdr.org>). RAST uses “subsystems technology”, where a subsystem is defined as a set of abstract functional roles. Subsystems range from single metabolic pathways, such as histidine degradation, or components of a cellular structure, such as SSU rRNA.

The steps RAST utilizes are identifying rRNA and tRNA genes, and protein-encoding genes, as well as gene function to provide an idea of the metabolic network present in a genome (Aziz et al., 2008).

The first step RAST uses to annotated genomes is to identify tRNA and rRNA genes. tRNA genes are found using tRNAscan (Lowe and Eddy, 1997), and rRNA with “search_for_rnas” (Aziz et al., 2008). Protein-encoding genes are found using GLIMMER, which is a computational gene finder that uses an Interpolated Markov Model and provides an estimate of probable genes (Delcher et al., 1999). GLIMMER is able to identify approximately 97-98% of all genes, but most of the genes missed by GLIMMER appear to encode hypothetical proteins whose existence is only supported by other programs (Delcher et al. 1999). Therefore, GLIMMER and RAST may not be the most useful annotation system for genomes of uncharacterized organisms.

Next, a set of representative sequences from a set of FIGfams are put together (determined by subsystems). FIGfams are another type of protein family, and each family contains a set of proteins, a family function and a decision procedure. The proteins in a family are homologous and share a common function, and a decision procedure uses an input protein sequence and returns a decision about whether the protein is globally similar and should be added to the family or not. The characteristic of the representative sequences selected from a set of FIGfams is that they are universal (or nearly universal) in prokaryotes. These FIGfams are compared to the protein-encoding genes of the genome being annotated, to provide a small number of genes used to estimate the closest phylogenetic neighbours (Aziz et al. 2008). Phylogenetic neighbours are then used to find FIGfams likely to be present in the genome being annotated, and the FIGfams are

searched for in the new genome. When genes are found, they become “determined genes” and are no longer considered “putative genes” for the remaining part of the annotation process. The set of determined genes at this point is used as a training set to look for more genes among the remaining putative genes. Remaining putative genes are compared to the entire collection of FIGfams. After the use of FIGfams, the RAST server will look for sequence similarity-based assignment of function for the remaining putative genes using a BLAST search against a large non-redundant protein database. The set of determined genes is then used for metabolic reconstruction, or connecting genes to functions (Aziz et al. 2008).

1.2.1.2 JGI-MGAP

The United States Department of Energy runs the Joint Genome Institute (JGI) Microbial Genome Annotation Pipeline (MGAP) (<http://img.jgi.doe.gov>), which performs structural and functional annotation of microbial genomes. MGAP has three major steps for microbial annotation (Huntemann et al., 2015). First, the sequenced data is preprocessed to ensure high quality data for annotation which includes removing sequences shorter than 150 nucleotides, trimming sequences, low complexity filtering, and for finished circular genomes the origin of replication is found. Second, the sequence data is analyzed to predict where genes occur (gene calling) for structural annotation. Finally, the genes found in the structural annotation are compared to protein databases to provide a functional annotation.

Structural annotation is performed in MGAP by detection of CRISPR arrays, detection of non-protein coding RNA genes and the prediction of protein coding genes. Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) are a type of direct repeat in prokaryotic DNA,

with unique spacer sequences between the repeats. CRISPRs have a unique function of protecting their host against invading extrachromosomal elements, such as bacteriophages, plasmids, and viruses (Barrangou et al., 2007; Mojica et al., 2005). In MGAP, CRISPRs are identified using two programs called CRT and PILER-CR (Huntemann et al., 2015). Both protein-coding and non-coding RNA genes are identified using Hidden Markov Models and sequence similarity-based approaches in MGAP. Some of the programs MGAP uses to find non-protein coding RNAs are tRNAscan (tRNAs), hmmsearch (ribosomal RNA genes) and their own Hidden Markov Models (ribosomal RNA genes) (Huntemann et al., 2015). Protein-coding genes are found in MGAP using a gene prediction program called Prodigal, where Prodigal calculates log-likelihood functions rather than the often used Hidden Markov Models (Hyatt et al., 2010). After this step in MGAP, each annotated gene receives a locus tag, and a GenBank format genome data file is created.

After the structural annotation, the GenBank files give protein-coding genes that are compared to protein families and proteomes to provide gene product names or functions (Huntemann et al., 2015). A combination of COG, KEGG, MetaCyc, Pfam and TIGRfam are used in MGAP to give the intended functional annotation and provide protein product names. COGs are groups of proteins that are classified by function (Tatusov et al., 2001). KEGG, or the Kyoto Encyclopedia of Genes and Genomes, is a collection of manually curated orthologous groups which provide information about an organism's metabolism and phenotype (Kanehisa et al., 2014). KEGG assignment is associated with MetaCyc reactions. MetaCyc is a universal database of metabolic pathways and enzymes that is curated from scientific literature, and therefore contains experimentally derived knowledge (Caspi et al., 2008). Pfam and TIGRfam are two more types of protein families used for comparison. Pfam are protein families which are represented by multiple

sequence alignments and Hidden Markov Models (Finn et al., 2014). TIGRfam is a collection of manually curated protein families. Hidden Markov Models, multiple sequence alignments, Gene Ontology assignments, literature references, and Pfam models all contribute to TIGRfam protein family classification (Haft et al., 2003).

1.2.2 Applications in microbiology

Comparative genomics allows insight into features that distinguish pathogens and non-pathogens, bacterial physiology and bacterial evolution. Investigations often involve identifying genes shared between genomes, or unique to a genome. This refers to what are considered the “core” and “accessory” genomes. The core genome contains the genes of an organism that are shared among all members of that species, and the accessory genome contains elements that are present in some strains, but absent in others. Comparative genomics can also utilize comparing specific genes of interest, which can include virulence genes. One study of endophytes (bacteria living within plant tissue), investigated the differences of known virulence genes between endophytes and other symbiotic bacteria (Lopez-Fernandez et al., 2015). Differences between the strains were related to divergences in metabolic pathways, and led to the identification of features that distinguish harmful and harmless bacteria. Another example of the application of comparative genomics approaches to discriminating pathogenic and non-pathogenic bacteria is a study of *Listeria monocytogenes* (a food-borne pathogen) and non-pathogenic *L. innocua* in which predicted genes involved in surface and secreted proteins, transporters, and transcriptional regulators unique to each strain were identified (Glaser et al., 2001).

While comparative genomics is a useful tool to identify differences among bacteria, there are limitations to this approach. For example, low quality sequence data or low coverage of genome sequences could simply result in genes being missed during annotation. While genomes are becoming more readily available for different organisms, variation within species still exists. Analysis with few genomes may identify differences that are due to variation, and are “false positives”, because there is not sufficient representation of the species. Most significantly, differences between genomes may correspond to ORFs of unknown function. It becomes difficult to understand the significance of the differences when functions are unknown. It can be especially difficult when working with organisms that are not well characterized, such as *Brachyspira*, and follow-up laboratory experiments are key. An example of experimental follow-up is a study where *Klebsiella pneumoniae* strains of differing pathogenicity were compared to identify putative virulence factors. Four genomic islands were identified, which included genes for secretion systems (Lery et al., 2014). To follow their comparative genomics analysis, a mouse inoculation trial was performed with type VI secretion system mutants and the mutant was avirulent. The gene was expressed *in vivo* and associated with *K. pneumoniae* isolates from severe infections, suggesting its importance in the virulence of *K. pneumoniae* (Lery et al., 2014). The experiments following the original comparative genomics analysis were instrumental in demonstrating the importance of the differences to pathogenicity.

2 Objectives

The re-emergence of mucohaemorrhagic diarrhea in pigs associated with *Brachyspira* is a challenge for swine producers. Limited understanding of *Brachyspira* pathogenesis and virulence factors causing the disease, make it difficult to treat or prevent infection. Insight into pathogen biology and the development of new diagnostic tools are key to controlling *Brachyspira*-associated infections. To develop understanding in these areas, the following objectives were addressed in this thesis:

1. To identify putative virulence factors of “*B. hampsonii*” by comparing the genomes of pathogenic “*B. hampsonii*” strain 30446 and the non-pathogenic “*B. hampsonii*” strain KL-180.
2. To develop a diagnostic method for simultaneous detection of multiple *Brachyspira* species using deep sequencing of the *nox* gene target, and to use this method to determine the frequency of occurrence of multiple *Brachyspira* species in clinical cases of *Brachyspira*-associated disease.

3 Comparative genomics of pathogenic and non-pathogenic strains of “*Brachyspira hampsonii*” and relationship of gene content differences to putative virulence factors

3.1 Abstract

Brachyspira-associated disease results in mucohaemorrhagic diarrhea and colitis in pigs. The disease has re-emerged in North America and the re-emergence has been associated with a novel species, “*B. hampsonii*”. Through inoculation trials, “*B. hampsonii*” strains 30446 and 30599 have been demonstrated to be pathogenic in swine. “*B. hampsonii*” strains were also isolated from Lesser Snow Geese in the Canadian Arctic, where one particular strain had genetic similarity to “*B. hampsonii*” strain 30446. Despite the genetic similarity, strain KL-180 is non-pathogenic in swine. Whole genome sequences are available for “*B. hampsonii*” strain 30446 and KL-180, providing an opportunity to identify genomic determinants of virulence. “*B. hampsonii*” strain 30599 and *B. hyodysenteriae* strain WA1 were included as two additional pathogenic representatives. Genome annotations were performed using the RAST (Rapid Annotation using Subsystem Technology) server and the JGI Microbial Genome Annotation Pipeline (MGAP). The annotation obtained in RAST identified 25 genes unique to strain 30446 and 11 unique to KL-180. The JGI annotation provided 22 genes present in the pathogenic strains that were absent in the non-pathogenic strain KL-180, and this list contains 11 hypothetical proteins. Present in the pathogenic strains (and identified by both RAST and JGI) were a set of genes corresponding to the Streptolysin S (*sag*) operon. Streptolysin S is a virulence factor first characterized in *Streptococcus*, that contributes to hemolysis of the host blood cells. The genes of this operon were found bioinformatically and biologically through PCR, and were found in pathogenic “*B. hampsonii*” strains. Screening for the genes in *Brachyspira* ATCC strains revealed their presence in *B. innocens*, *B. murdochii* and *B. intermedia* which are not considered to be pathogens. The genes are present, but it is not known whether they are expressed or functional in these non-

pathogenic strains. The presence of this operon in some pathogenic *Brachyspira* species, and absence in the non-pathogenic “*B. hampsonii*” strain KL-180 supports the possibility of the operon being a virulence factor and contributing to disease outcome.

3.2 Background

Brachyspira is associated with diarrhea and colitis in pigs, and the most severe disease (mucohaemorrhagic diarrhea) is traditionally associated with the species *B. hyodysenteriae* (Harris et al., 1972). Disease has re-emerged in North America, and is often associated with the novel species “*B. hampsonii*” (Chander et al., 2012; Harding et al., 2010). “*B. hampsonii*” has been isolated from pigs in North America (Burrough et al., 2012a; Harding et al., 2010) and Europe (Mahu et al., 2014; Rohde et al., 2014). “*B. hampsonii*” has been isolated from waterfowl in Spain (Martinez-Lobo et al., 2013), mallards in Sweden (Rasback et al., 2007) and Lesser Snow Geese in the Canadian Arctic (Rubin et al., 2013b). It is not unusual for *Brachyspira* spp. to be found in wildlife reservoirs, as there have been previous reports of *Brachyspira* spp. in bird populations (Jansson et al., 2004; Jansson et al., 2015; Oxberry et al., 1998). Birds, especially those that are migratory, have the ability to disperse bacteria over long distances and could play a role in *Brachyspira* ecology. Strains that were isolated from Snow Geese in the Arctic that belong to the novel species “*B. hampsonii*” are of particular interest (Rubin et al., 2013b).

Snow goose isolates were identified as belonging to the “*B. hampsonii*” species based on multi-locus sequence typing, *nox* and *cpn60* sequence data. One snow goose isolate, KL-180, was found to be 99% identical based on the *nox* gene to “*B. hampsonii*” clade II strain 30446, which has been shown to be pathogenic in pigs (Rubin et al., 2013a). DNA sequence identities for four

multi-locus sequence typing (MLST) targets (*est*, *glp*, *pgm*, and *thi*) and universal sequence target *cpn60* also demonstrated similarity between the two strains with over 95% pairwise sequence identity in each gene. Despite the genetic similarity to the pathogenic *Brachyspira* species, when inoculated into pigs, the KL-180 snow goose strain was not associated with clinical disease (Rubin et al., 2013b).

The pathogenesis of swine dysentery caused by *Brachyspira* is not completely understood. In infected animals, *Brachyspira* is observed in the crypts of the colon and putative virulence factors include motility, lipopolysaccharides and haemolysins (Bellgard et al., 2009; Naresh and Hampson, 2010; Nuessen et al., 1982). Since the publication of the original description of KL-180, its whole genome sequence has been determined. The availability of whole genome sequences for pathogenic and non-pathogenic isolates that are apparently closely related presents an opportunity to identify genomic determinants of virulence. The objective of this study was to use a comparative genomics approach to identify putative virulence factors of “*B. hampsonii*”.

3.3 Methods

3.3.1 *Brachyspira* genome sequences

The *Brachyspira* genome sequences used in this study are described in Table 3.1. “*B. hampsonii*” strain 30446 and “*B. hampsonii*” strain 30599 were isolated from colon contents or feces of pigs with mucohaemorrhagic diarrhea, and have been demonstrated through inoculation trials to be pathogenic in pigs (Costa et al., 2014b; Rubin et al., 2013a). *B. hyodysenteriae* strain WA1 is another pathogenic *Brachyspira* isolate (Bellgard et al., 2009), included along with “*B.*

hampsonii” strains as another pathogenic representative. “*B. hampsonii*” strain KL-180 was isolated from a Lesser Snow Goose sampled in the Canadian Arctic, and has been previously demonstrated to be non-pathogenic in pigs (Rubin et al., 2013b).

3.3.2 Genome annotation

Genome annotations of the *Brachyspira* species listed in Table 3.1 were carried out using the SEED-based automated annotation system provided by the RAST (Rapid Annotation using Subsystem Technology) server (Aziz et al., 2008), and the JGI Microbial Genome Annotation Pipeline (MGAP) (Huntemann et al., 2015). The MGAP sequence data processing consists of structural annotation by including identification of protein-coding genes, non-protein coding RNAs and regulatory RNA features (such as riboswitches), and CRISPR elements. Structural annotation is followed by the assignment of protein names and functions. The RAST server identifies protein coding, rRNA and tRNA genes, assigns functions to the genes and predicts functional subsystems that are present.

Genbank files of the annotated genomes (from JGI MGAP annotation) were uploaded to the GView server (<http://server.gview.ca>) to visualize genome comparisons through the pangenome analysis (Petkau et al., 2010). The pangenome analysis represents the full complement of genes present in all the genomes included, highlighting genes that are absent or present in certain genomes. Overall sequence identity between *Brachyspira* genomes was calculated as average nucleotide identity by MUMmer (ANIm) using JSpecies (v 1.2.1) (Richter and Rossello-Mora, 2009).

Table 3.1. Characteristics of genome sequences used in this study.

Species	Strain	Source	Pathogenicity in pigs	Size (Mb)	Assembly	BioProject Accession	Reference
<i>B. hyodysenteriae</i>	WA1	Pig	Pathogenic	3.04	Complete	PRJNA31263	(Bellgard et al., 2009)
<i>“B. hampsonii”</i>	30599	Pig	Pathogenic	2.94	612 contigs	PRJNA187424	(Costa et al., 2014b)
<i>“B. hampsonii”</i>	30446	Pig	Pathogenic	3.04	4 scaffolds	PRJNA169353	(Rubin et al., 2013a)
<i>“B. hampsonii”</i>	KL-180	Lesser Snow Goose	Non- pathogenic	3.09	6 scaffolds	PRJNA243522	(Chaban et al. Unpublished)

3.3.3 *Brachyspira* isolates and culture

“*B. hampsonii*” clade I isolate D11-30599 and clade II isolate D09-30446 were the first Canadian isolates of each clade derived from clinical cases of mucohaemorrhagic diarrhea. A collection of 443 “*B. hampsonii*” clinical isolates from our laboratory culture collection was reduced to epidemiologically distinct isolates, or isolates from different production systems, resulting in 9 representative isolates (Clade I: D11-30599, and D12-32613; Clade II: D09-30446, D11-29565-A6P8, D12-10616-P8, D12-27447, D12-27451, D12-32539, and D13-06010). Along with “*B. hampsonii*” strains, other *Brachyspira* ATCC strains were included (*B. pilosicoli* ATCC 51139^T, *B. murdochii* ATCC 51284^T, *B. innocens* ATCC 29796^T, and *B. hyodysenteriae* ATCC 27164^T). Long-term storage of *Brachyspira* isolates was at -80°C in brain-heart infusion (BHI) broth containing 10% (v/v) glycerol. Strains from storage were cultured in JBS broth, which consists of BHI with 1% (w/v) glucose, 5% (v/v) fetal bovine serum, 5% (v/v) sheep blood, and the cultures were incubated anaerobically at 42°C with stirring.

3.3.4 DNA extraction and PCR

Genomic DNA was purified from JBS cultures using a modified salting-out procedure, as previously described (Martin-Platero et al., 2007). Primers for PCR were designed using Primer3 (Koressaar and Remm, 2007; Untergasser et al., 2012) for genes in the *sag*-like operon found in “*Brachyspira hampsonii*” strains 30446 and 30599, as shown in Table 3.2. Genus-specific primers for the *nox* target were used as previously described (Rohde et al., 2002), and included as a positive control to confirm suitability of genomic DNA preparations for PCR.

Each PCR reaction for assays designed for the *sag*-like genes contained 1 x PCR reaction buffer (0.2 M Tris-HCl at pH 8.4, 0.5 M KCl), 2.5 mM MgCl₂, 200 µM dNTP mixture, 400 nM each primer, 2.5 U Taq DNA Polymerase and 2µl of template DNA, in a final volume of 50 µl. PCR was performed in a thermocycler (Eppendorf Mastercycler) with the following program: 94°C for 5 min, 40 cycles of 95°C for 30 s, 60°C for 30 s, 72°C for 30 s, followed by a final extension of 72°C for 2 min. PCR products were visualized on a 1% agarose gel by ethidium bromide staining.

Table 3.2. Primer sequences used in this study.

Target	Primer Name	Primer sequence (5'-3')	Amplicon size (bp)	Reference
<i>sagA</i>	JH0736	TAA ACT AMG AAC AAC ATG TTC TCT AAA	100	This study
	JH0737	TGT TAC GCT AAM GCA GCA WG		
<i>sagB</i>	JH0738	TGT TTT CGG AGG CGG TAT AG	161	This study
	JH0739	TTG CAG CAG CAG TAA GCT GT		
<i>sagC</i>	JH0740	CCG GAT GTT ATG AAT GTT TTG A	168	This study
	JH0741	AAA ATG CTT CAT ATA ATG CAA GAG AA		
<i>sagD</i>	JH0742	TTG CGG TCA TCA GAC AGG TA	209	This study
	JH0743	TTT CAA TGC CTT CAC CCA AT		
<i>sagE</i>	JH0744	CTG CTT GCC ATT TCT TTT CC	239	This study
	JH0745	AAA TGA ATC AGC ATT GAC TCA AAT A		
<i>sagG</i>	JH0746	TGG TTC RGG AAA AAC KAC ATT	250	This study
	JH0747	AGC AAG AGT TTT CGG ATT CAA		
<i>sagI</i>	JH0748	TGC TTA AAG TAA CGC CTC TTC C	173	This study
	JH0749	TGC CAA TAT AAT TAC AGG CAA AAA		
<i>nox</i>	noxF	TAG CYT GCG GTA TYG CWC TTT GG	939	(Rohde et al., 2002)
	noxR	CTT CAG ACC AYC CAG TAG AAG CC		

3.4 Results

3.4.1 Genome annotation

Brachyspira genome sequences were annotated using both the RAST server and the JGI Microbial Genome Annotation Pipeline. Two separate annotations were performed to determine whether either method provided different genes in comparative analyses. The overall characteristics of the annotation by each method are described in Table 3.3, where they resulted in similar annotations. Each genome was approximately 3 Mb in size, and had a GC content of 27%. From the RAST annotation, the four genomes had an average of 305 subsystems present. A subsystem is defined as a collection of functional roles, which together implement a specific biological process or structural complex or perform related functional roles (Overbeek et al., 2005). Subsystems may comprise single metabolic pathways, such as histidine degradation, or components of a cellular structure, such as SSU rRNA. The number of coding sequences ranged from 2644 in *B. hyodysenteriae* WA1 to 2965 in “*B. hampsonii*” strain 30599, and 36 RNA genes were found in each of the four genomes with the RAST annotation. RAST does not report tRNA and rRNA genes separately, rather together as RNA genes. With the JGI annotation, the number of coding sequences ranged from 2572 in “*B. hampsonii*” strain 30446 to 2736 in “*B. hampsonii*” strain 30599. Thirty-four RNA genes were detected in both “*B. hampsonii*” strain 30446 and strain 30599, (33 tRNA genes and 1 rRNA gene), 37 RNA genes in *B. hyodysenteriae* WA1 (34 tRNA genes and 3 rRNA genes), and 36 RNA genes in “*B. hampsonii*” strain KL-180 (33 tRNA genes and 3 rRNA genes).

The pangenome analysis from the GView server illustrated the similarity of the four *Brachyspira* genomes (Figure 3.1). The pangenome represents the full complement of genes present in all the genomes. Individual genomes are then mapped on to the pangenome to highlight genes that are absent or present in individual genomes. The innermost ring (light green) represents the combination of all the genomes (“pangenome”). The outer rings show predicted genes that are present in the respective genomes and gaps indicate the gene was not detected, highlighting gene content differences between the *Brachyspira* genomes. The pattern that is most biologically interesting in this study is highlighted in the figure: genes are present in the pathogens (*B. hyodysenteriae* and “*B. hampsonii*” strain 30446 and strain 30599), and absent in the non-pathogen (“*B. hampsonii*” strain KL-180).

Overall sequence identity between *Brachyspira* genomes was calculated as average nucleotide identity by MUMmer (ANI_m). Generally, an average nucleotide percent identity over 95% in ANI_m is indicative of the same species (Richter and Rossello-Mora, 2009). By that criterion, strains “*B. hampsonii*” strain 30446 and “*B. hampsonii*” strain KL-180 can be considered the same species with an ANI_m of 97.34%. (Table 3.4). “*B. hampsonii*” strain 30599 differs from both “*B. hampsonii*” strain 30446 and “*B. hampsonii*” strain KL-180 with ANI_m values of 93.51% and 93.58%, respectively. Though all strains belong to “*B. hampsonii*”, the differences in ANI_m are not surprising as “*B. hampsonii*” strains 30446 and KL-180 belong to clade II and “*B. hampsonii*” strain 30599 belongs to clade I. All the “*B. hampsonii*” strains differed from the other species, *B. hyodysenteriae*, with an average ANI_m of 88.87%.

Table 3.3. Genome annotation characteristics obtained from (A) the RAST server, and (B) the JGI MGAP.

(A)

		<i>B. hyodysenteriae</i> strain WA1	<i>“B. hampsonii”</i> strain 30446	<i>“B. hampsonii”</i> strain 30599	<i>“B. hampsonii”</i> strain KL-180
RAST	Size (bp)	3036634	3019251	2943158	3093592
	GC content (%)	27.0	27.4	27.8	27.5
	# of subsystems	308	311	295	306
	Coding sequences	2644	2655	2965	2706
	rRNA and tRNA genes	36	36	36	36

(B)

		<i>B. hyodysenteriae</i> strain WA1	<i>“B. hampsonii”</i> strain 30446	<i>“B. hampsonii”</i> strain 30599	<i>“B. hampsonii”</i> strain KL-180
JGI-MGAP	Size (bp)	3036634	3019251	2943158	3093592
	GC content (%)	27.0	27.4	27.8	27.5
	Coding Sequences	2642	2572	2736	2706
	Genes with predicted protein function	1808	1872	1927	1849
	rRNA genes	3	1	1	3
	tRNA genes	34	33	33	33

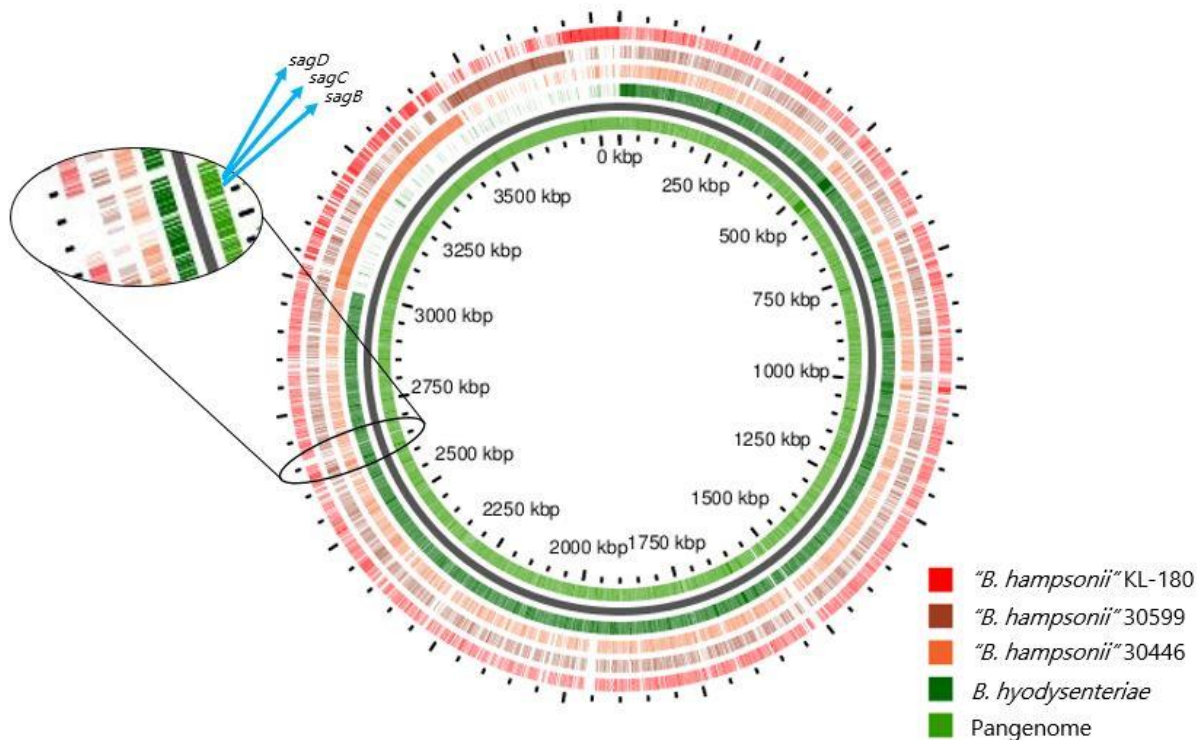


Figure 3.1. Pangenome analysis obtained from the GView server.

Genbank formatted files containing annotation information were obtained from the JGI IMG, and uploaded to the Gview server (<http://gview.server.ca>) for a pangenome analysis (Petkau et al., 2010). The innermost ring (light green) represents the combination of all the genomes (“pangenome”). The outer rings show predicted genes that present in the respective genomes and gaps indicate the gene was not detected, highlighting gene content differences between the *Brachyspira* genomes. Highlighted is the pattern that is most biologically interesting in this study: a combination where genes are present in the pathogens (*B. hyodysenteriae* and “*B. hampsonii*” strain 30446 and strain 30599), and absent in the non-pathogen (“*B. hampsonii*” strain KL-180). The highlighted region also corresponded to a set of genes that is similar to the Streptolysin S (*sag*) operon found in other bacteria.

Table 3.4. Average Nucleotide Identity (ANI_m) values for pairwise comparisons between four *Brachyspira* spp..

	<i>“B. hampsonii”</i> strain 30446	<i>“B. hampsonii”</i> strain 30599	<i>“B. hampsonii”</i> strain KL-180	<i>B. hyodysenteriae</i> WA1
<i>“B. hampsonii”</i> strain 30446	--	93.51	97.34	88.69
<i>“B. hampsonii”</i> strain 30599		--	93.58	89.15
<i>“B. hampsonii”</i> strain KL-180			--	88.77
<i>B. hyodysenteriae</i> WA1				--

3.4.2 Clusters of Orthologous Groups (COGs)

Clusters of Orthologous Groups (COGs) are categories of proteins that are related by their functions (Tatusov et al., 2001). The distributions of functional genes in different COG categories for each *Brachyspira* genome were compared, with COG representation based on the JGI MGAP annotation. The four *Brachyspira* genomes had similar predicted functional repertoires, with “*B. hampsonii*” strain 30599 containing the lowest number of COGs in most functional categories (Figure 3.2). The functional categories with the largest proportion of COGs in the four *Brachyspira* genomes are E (amino acid transport and metabolism), G (carbohydrate transport and metabolism), J (translation, ribosomal structure, and biogenesis) and R (general function prediction only). Functional category G (carbohydrate transport and metabolism) was the only category where “*B. hampsonii*” strain KL-180 contained fewer COGs than the pathogenic strains. The proportion of proteins found in the functional COG categories ranged from 51 to 59% in the four *Brachyspira* proteomes, with 41 to 49% not identified within the COG categories.

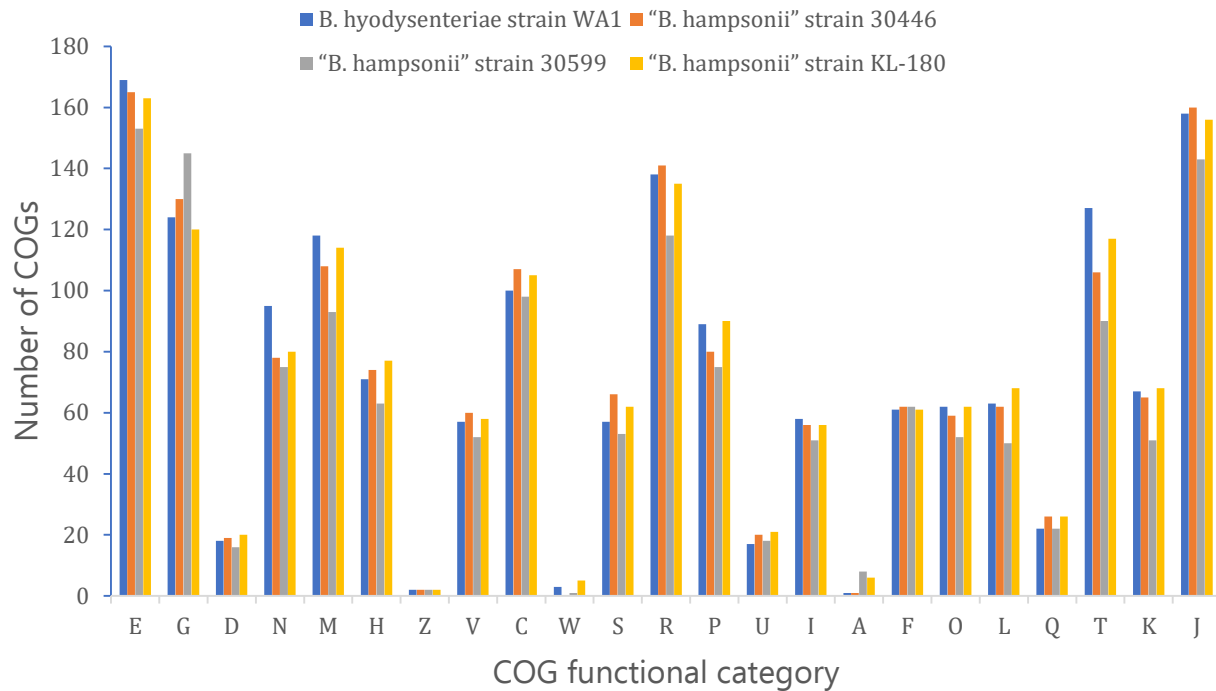


Figure 3.2. Distribution of COGs in *Brachyspira* genomes.

COG categories: E: Amino acid transport and metabolism; G: Carbohydrate transport and metabolism; D: Cell cycle control, cell division, chromosome partitioning; N: Cell motility; M: Cell wall/membrane/envelope biogenesis; H: Coenzyme transport and metabolism; Z: Cytoskeleton; V: Defense mechanisms; C: Energy production and conversion; W: Extracellular structures; S: Function unknown; R: General function prediction only; P: Inorganic ion transport and metabolism; U: Intracellular trafficking, secretion, and vesicular transport; I: Lipid transport and metabolism; A: RNA processing and modification; F: Nucleotide transport and metabolism; O: Post-translational modification, protein turnover, and chaperones; L: Replication, recombination and repair; Q: Secondary metabolites; T: Signal transduction mechanisms; K: Transcription; J: Translation, ribosomal structure and biogenesis.

3.4.3 Functional comparison of “*B. hampsonii*” strains 30446 and KL-180 in RAST

Using the SEED viewer from the RAST genome annotation, a functional comparison was performed between the pathogenic “*B. hampsonii*” strain 30446 and the non-pathogenic “*B. hampsonii*” strain KL-180 (Tables 3.5 and 3.6). The functional comparison generated a list of 25 genes, that are associated with a subsystem and that are present in strain 30446 and absent in strain KL-180 (Table 3.5). These 25 genes are distributed among 9 different broad categories, with the highest number of genes being found in the virulence, disease and defense category, followed by the membrane transport and carbohydrate metabolism categories.

One operon present in “*B. hampsonii*” strain 30446 and absent in “*B. hampsonii*” strain KL-180 was genes *nikA-E*, which are genes involved in nickel transport (Table 3.5). Nickel is an essential cofactor for a number of enzymatic reactions including production and consumption of hydrogen, hydrolysis of urea, oxidation of carbon monoxide, and detoxification of superoxide anion radicals (Eitinger and Mandrand-Berthelot, 2000). Another operon found in strain 30446 but absent in KL-180 is the *sag* operon, which is responsible for the production Streptolysin S (SLS). SLS is a cytolytic toxin and virulence factor in *Streptococcus* spp. (Molloy et al., 2011).

The functional comparison generated a list of 11 genes, that are associated with a subsystem and that are present in strain KL-180 and absent in strain 30446 (Table 3.6). The genes unique to “*B. hampsonii*” strain KL-180, or the non-pathogenic strain, were all from different subsystems and did not form any operons. Although any of the genes found, whether annotated or uncharacterized, in either only the pathogenic or non-pathogenic strain could be important in disease development, the remainder of this study focused on investigating the set of genes found

in “*B. hampsonii*” strain 30446 that corresponded to genes similar to the Streptolysin S (*sag*) operon. This set of genes was of obvious interest because the *sag* operon is a virulence factor of *Streptococcus* spp. where it acts as a hemolysin, breaking down erythrocytes of its host (Molloy et al., 2011), and haemolysis appears to be a virulence factor in *Brachyspira*-associated disease.

Table 3.5. Genes associated with a subsystem in “*B. hampsonii*” strain 30446, and absent in “*B. hampsonii*” strain KL-180, based on the metabolic construction in RAST.

Gene	Subsystem in RAST	Functional Category
N-Ac+E2:E26etyl-D-glucosamine ABC transport system, permease protein 1	Chitin and N-acetylglucosamine utilization	Carbohydrates
Phosphoglycerate kinase (EC 2.7.2.3)	Entner-Doudoroff Pathway	Carbohydrates
PTS system, sucrose-specific IIA component (EC 2.7.1.69)	Sucrose utilization	Carbohydrates
Stage 0 sporulation protein J	Bacterial Cytoskeleton	Cell Division and Cell Cycle
Bipolar DNA helicase HerA	Rad50-Mre11 DNA repair cluster	Clustering-based subsystems
D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4)	CBSS-84588.1.peg.1247	Clustering-based subsystems
TPR domain protein in aerotolerance operon	Aerotolerance operon in Bacteroides and potentially orthologous operons in other organisms	Clustering-based subsystems
tRNA pseudouridine synthase B (EC 4.2.1.70)	CBSS-138119.3.peg.2719	Clustering-based subsystems
DNA recombination and repair protein RecO	DNA repair, bacterial RecFOR pathway	DNA Metabolism
Nickel ABC transporter, periplasmic nickel-binding protein NikA (TC 3.A.1.5.3)	Transport of Nickel and Cobalt	Membrane Transport
Nickel transport ATP-binding protein NikD (TC 3.A.1.5.3)	Transport of Nickel and Cobalt	Membrane Transport
Nickel transport ATP-binding protein NikE (TC 3.A.1.5.3)	Transport of Nickel and Cobalt	Membrane Transport
Nickel transport system permease protein NikB (TC 3.A.1.5.3)	Transport of Nickel and Cobalt	Membrane Transport
Nickel transport system permease protein NikC (TC 3.A.1.5.3)	Transport of Nickel and Cobalt	Membrane Transport

Thymidine kinase (EC 2.7.1.21)	pyrimidine conversions	Nucleosides and Nucleotides
Pseudouridine 5'-phosphate glycosidase	Pseudouridine catabolism	Nucleosides and Nucleotides
Pseudouridine kinase (EC 2.7.1.83)	Pseudouridine catabolism	Nucleosides and Nucleotides
RelB/StbD replicon stabilization protein (antitoxin to RelE/StbE)	Toxin-antitoxin replicon stabilization systems	Regulation and Cell signaling
Streptolysin S export protein (SagG)	Thiazole- oxazole-modified microcin (TOMM) synthesis	Secondary Metabolism
Export ABC transporter ATP-binding protein	Streptolysin S Biosynthesis and Transport	Virulence, Disease and Defense
Streptolysin S biosynthesis protein B (SagB)	Streptolysin S Biosynthesis and Transport	Virulence, Disease and Defense
Streptolysin S biosynthesis protein C (SagC)	Streptolysin S Biosynthesis and Transport	Virulence, Disease and Defense
Streptolysin S biosynthesis protein D (SagD)	Streptolysin S Biosynthesis and Transport	Virulence, Disease and Defense
Streptolysin S export transmembrane permease (SagI)	Streptolysin S Biosynthesis and Transport	Virulence, Disease and Defense
Streptolysin S self-immunity protein (SagE)	Streptolysin S Biosynthesis and Transport	Virulence, Disease and Defense

Table 3.6. Genes associated with a subsystem in “*B. hampsonii*” strain KL-180, and absent in “*B. hampsonii*” strain 30446, based on the metabolic construction in RAST.

Gene	Subsystem	Functional Category
D-serine dehydratase (EC 4.3.1.18)	Glycine and Serine Utilization	Amino Acids and Derivatives
N-Acetylneuraminate cytidyltransferase (EC 2.7.7.43)	CMP-N-acetylneuraminate Biosynthesis	Cell Wall and Capsule
LSU ribosomal protein L34p	Cell Division Subsystem including YidCD	Clustering-based subsystems
DNA double-strand break repair Rad50 ATPase	Rad50-Mre11 DNA repair cluster	Clustering-based subsystems
Histone acetyltransferase HPA2 and related acetyltransferases	CBSS-216591.1.peg.168	Clustering-based subsystems
Ribonucleotide reductase of class III (anaerobic), activating protein (EC 1.97.1.4)	Ribonucleotide reduction	Nucleosides and Nucleotides
Phage tail length tape-measure protein	Phage tail proteins	Phages, Prophages, Transposable Elements
ATP-dependent Clp protease ATP-binding subunit ClpA	Proteolysis in bacteria, ATP-dependent	Protein Metabolism
RNA-2'-3'-PO4:RNA-5'-OH ligase	RNA 3'-terminal phosphate cyclase	RNA Metabolism
Anti-sigma B factor antagonist RsbV	SigmaB stress response regulation	Stress Response
ABC-type nitrate/sulfonate/bicarbonate transport system, permease component	Alkanesulfonate assimilation	Sulfur Metabolism

3.4.4 *sag*-like genes in *Brachyspira*

The *sag* operon is responsible for the production Streptolysin S (SLS), which is a cytolytic toxin and virulence factor in *Streptococcus* spp. (Molloy et al., 2011). In *Streptococcus*, the operon comprises a contiguous set of nine genes called *sagA* through *sagI*. The RAST annotation identified genes in the *sag*-like operon in “*B. hampsonii*” strain 30446, except for *sagA* and *sagF* (Table 3.3). These observations were partially confirmed by the JGI annotation, which identified *sagB*, *sagD*, an amino terminal protease, a hypothetical protein and an ABC transporter protein all in the same region (Table 3.7). While the RAST comparison described above was only between strain 30446 and KL-180, the JGI utilized a comparison of all four genomes to identify proteins present in the three pathogenic *Brachyspira* species included and absent in non-pathogenic “*B. hampsonii*” KL-180. This comparison aided in identifying “universal” traits of pathogenic *Brachyspira*, rather than traits found in one pathogen or another. The JGI annotation also identified 11 hypothetical proteins unique to the three pathogenic *Brachyspira* strains (Table 3.7), including the *sag* operon genes described above.

Table 3.7. Genes present in *B. hyodysenteriae* strain WA1, “*B. hampsonii*” strain 30446 and strain 30599, but absent in “*B. hampsonii*” strain KL-180, as determined from annotation on JGI MGAP.

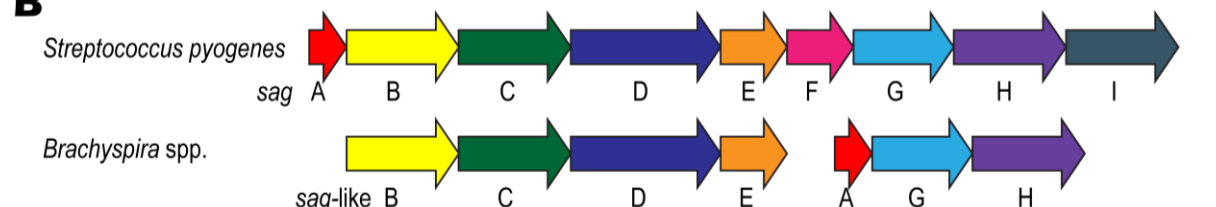
Locus Tag in <i>B. hyodysenteriae</i> WA1	Gene Name
BHWA1_00016	hypothetical protein
BHWA1_00128	hypothetical protein
BHWA1_00152	Plasmid encoded toxin Txe
BHWA1_00480	ankyrin repeat-containing protein
BHWA1_00653	hypothetical protein
BHWA1_00861	hypothetical protein
BHWA1_00902	putative ATP-binding protein
BHWA1_01091	hypothetical protein
BHWA1_01130	hypothetical protein
BHWA1_01587	GltP, Na ⁺ /H ⁺ -dicarboxylate symporters putative
BHWA1_01616	TPR domain-containing protein
BHWA1_01818	plasmid encoded toxin
BHWA1_01819	hypothetical protein
BHWA1_01849	hypothetical protein
BHWA1_01964	TPR domain-containing protein
BHWA1_02353	streptolysin associated protein SagB
BHWA1_02355	streptolysin associated protein SagD
BHWA1_02356	CAAX amino terminal protease family protein
BHWA1_02357	hypothetical protein
BHWA1_02359	ABC transporter integral membrane type-2 domain containing protein
BHWA1_02578	hypothetical protein
BHWA1_02579	hypothetical protein

3.4.5 Locating *sagA* in *Brachyspira* genomes

In *Streptococcus* spp. the *sag* operon is contiguous, however manual investigation of the sequence upstream of *sagB* in the genomes of “*B. hampsonii*” strains 30599 and 30446 did not uncover any sequence that resembled *sagA*. Authors of a previous study utilizing genome mining to identify ribosomally synthesized and post-translationally modified peptides in anaerobic bacteria, however, found *sag*-like genes in *B. intermedia* and *B. hyodysenteriae* with *sagA* in a noncanonical location (Letzel et al., 2014). In these species, the gene encoding the peptide precursor was found between *sagF* and *sagG*, rather than upstream of *sagB*. Using this knowledge, and the protein sequences of the *sagA* genes in other *Brachyspira* species for comparison, the *sagA* gene was identified in both “*B. hampsonii*” strain 30446 and “*B. hampsonii*” strain 30599. Genome sequences upstream of *sagG* were manually evaluated to locate the presence of the *sagA* gene. The sequence is shown in Figure 3.3, along with *sagA* sequences identified through genome mining (Letzel et al., 2014).

A

<i>sagA</i> (<i>Streptococcus pyogenes</i>)*	MLKFTSNILATSVAETTQVAPGGCCCCCTTCCFSIATGSGNSQGGSGSYTPGK
<i>B. intermedia</i> precursor*	MIKYSSKVRTTCSLKSOSVVLTPGGCCCCCCCCSCCVNVTITNN
<i>B. hyodysenteriae</i> precursor*	MKYSSKLSTTCSLKSQSVVLAPGGCCCCCCCCSCCISVTVSK
" <i>B. hampsonii</i> " 30446 precursor	MIKYSSKLRTTCSLKSQSVVLTPGGCCCCCCCCSCCFSVTTK
" <i>B. hampsonii</i> " 30599 precursor	MIRYSSKLRTTCSLKSQSVVLTPGACCCSCCCSCCVSVTVNN

B**Figure 3.3. *sag*-like operon in *Brachyspira*.**

(A) Precursor peptide sequence, *sagA* in *Streptococcus pyogenes* and *Brachyspira* spp. (B) Organization of Streptolysin S (*sag*) operon, and *sag*-like operon in *Brachyspira*.

*Denotes sequences published by Letzel et al., 2014.

3.4.6 *sag*-like genes in “*B. hampsonii*” and other *Brachyspira* spp.

To allow for the detection of the *sag*-like genes in other *Brachyspira* spp., PCR primers were designed for each gene based on the sequences of “*B. hampsonii*” strain 30446 and “*B. hampsonii*” strain 30599. The primers were used on strain 30446, and the non-pathogenic strain KL-180, and results confirmed the presence of all *sag*-like genes in strain 30446 and not in KL-180 (Table 3.8). The *sag*-like genes could have been missed in strain KL-180 through the annotation analysis if, because the genome sequence of strain KL-180 contains some gaps within assembled scaffolds, the genes were not present in the assembled genome sequence. The inability to detect through PCR further supports the absence of *sag*-like genes in the non-pathogenic “*B. hampsonii*” strain KL-180.

If the *sag* operon is a virulence factor in *Brachyspira* pathogenesis, it would be expected to be universally present in pathogenic species and strains. Epidemiologically distinct “*B. hampsonii*” strains were screened for the *sag*-like genes, including two strains from Clade I and seven from Clade II (Table 3.8). All the *sag*-like genes were present in strains D09-30446, D11-29565-A6P8, D12-10616-P8, D12-27447, D12-27451, D11-30599, and D12-32613. *sagA* was not detected in D12-32529 and D13-06010. In D13-06010, *sagG* and *sagI* were not detected. Type strains of *B. pilosicoli*, *B. murdochii*, *B. innocens* and *B. hyodysenteriae* were also screened for the *sag*-like genes (Table 3.8). All the genes were detected in *B. murdochii*, *B. innocens* and *B. hyodysenteriae*. In *B. pilosicoli*, *sagA* and *sagD* were not detected.

The *sag*-like sequences obtained for “*B. hampsonii*” strain 30446 were used as BLASTn queries to determine if these genes are present in other *Brachyspira* spp. Genome sequences are

available on NCBI for *B. pilosicoli* P43/6/78 (Bioproject Accession: PRJNA67875), *B. intermedia* PWS/A (Bioproject Accession: PRJNA68051), and *B. murdochii* DSM 12563 (Bioproject Accession: PRJNA29543). In the genome sequences available, BLASTn revealed the *sag*-like genes to be present in all three genomes.

Table 3.8. Detection of *sag*-like genes in *Brachyspira* spp. by conventional PCR.

	Isolate/Strain	<i>sagA</i>	<i>sagB</i>	<i>sagC</i>	<i>sagD</i>	<i>sagE</i>	<i>sagG</i>	<i>sagI</i>
<i>“B. hampsonii”</i> Clade II	D09-30446	+	+	+	+	+	+	+
	KL-180	-	-	-	-	-	-	-
	D11-29565-A6P8	+	+	+	+	+	+	+
	D12-10616-P8	+	+	+	+	+	+	+
	D12-27447	+	+	+	+	+	+	+
	D12-27451	+	+	+	+	+	+	+
	D12-32539	-	+	+	+	+	+	+
	D13-06010	-	+	+	+	+	-	-
<i>“B. hampsonii”</i> Clade I	D11-30599	+	+	+	+	+	+	+
	D12-32613	+	+	+	+	+	+	+
<i>B. pilosicoli</i>	ATCC 51139 ^T	-	+	+	-	+	+	+
<i>B. murdochii</i>	ATCC 51284 ^T	+	+	+	+	+	+	+
<i>B. innocens</i>	ATCC 29796 ^T	+	+	+	+	+	+	+
<i>B. hyodysenteriae</i>	ATCC 27164 ^T	+	+	+	+	+	+	+

3.5 Discussion

The pathogenesis of *Brachyspira*-associated disease in swine is not fully understood. A unique opportunity for comparative genomics of *Brachyspira* strains to identify putative virulence factors occurred due to the availability of whole genome sequences of pathogenic and non-pathogenic strains of *Brachyspira* with remarkable overall genetic similarity, despite their pathogenic differences.

The overall genome characteristics described in Table 3.1 (size, GC content, number of predicted open reading frames) demonstrate the overall similarity of strains 30446 and KL-180 and suggest that the genomic differences that contribute to determining virulence are relatively small. Average Nucleotide Identity values further demonstrated the genetic similarity between pathogenic “*B. hampsonii*” strain 30446 and non-pathogenic “*B. hampsonii*” strain KL-180. The ANI_m between the two was 97.34%, suggesting beyond *nox*, *cpn60* and MLST target sequencing that the two strains belong to the same species. Generally, ANI values over 95% are considered to belong to the same species (Richter and Rossello-Mora, 2009).

With the COG analysis, the proportion of proteins found in the functional COG categories ranged from 51 to 59% in the four *Brachyspira* proteomes, with 41 to 49% not identified within the COG categories. In a range of prokaryotic genomes, the range of COGs that can be identified comprise 60 to 86% of the proteome (Galperin et al., 2015). The assignment rate within these *Brachyspira* spp. is lower than what is expected with other prokaryotes, highlighting there are a lot of unknown genes and proteins in *Brachyspira* that are not well represented in the databases

used to generated COGs. “*B. hampsonii*” strain 30599 had the lowest number of genes assigned, which could be due to the relatively poor assembly of the genome (Table 3.1).

The pangenome analysis from the GView server, again highlighted the similarity of the two genomes. Pathogenic strain “*B. hampsonii*” strain 30599 and *B. hyodysenteriae* were included as two additional pathogenic representatives. The pangenome analysis highlighted regions where the pattern that is most biologically interesting in this study exists: a combination where genes are present in the pathogens (*B. hyodysenteriae* and “*B. hampsonii*” strain 30446 and strain 30599), and absent in the non-pathogen (“*B. hampsonii*” strain KL-180). The evolution to pathogenic from non-pathogenic ancestors has been described principally by the acquisition of virulence genes on plasmids and pathogenicity islands by horizontal gene transfer (Allewelt et al., 2000; Censini et al., 1996; Gal-Mor and Finlay, 2006; Nishibuchi and Kaper, 1995). Other work has demonstrated, however, that some epidemic pathogens (such as *Rickettsia* spp. and *Mycobacteria* spp.) contain fewer virulence factors than non-epidemic species, and have reduced genomes (Georgiades, 2012; Maurelli, 2007). A study of *Yersinia pestis* (the bacterial agent of the plague) demonstrated mutational loss of glycosidase activity may have led to pathogenicity of the agent, and evolution of flea-borne transmission (Erickson et al., 2008). Thus, the possibility that virulence could be the result of the *absence* of a gene or genome region cannot be ruled out.

From the genome annotations of *Brachyspira* spp., genes unique to individual genomes were identified. From the annotation obtained in RAST, comparison between “*B. hampsonii*” strain 30446 (pathogenic) and “*B. hampsonii*” strain KL-180 (non-pathogenic) identified 25 genes unique to strain 30446 and 11 unique to KL-180. Any of the genes in these lists could be important

in *Brachyspira*-associated disease pathogenesis, either as a virulence determinant or a gene whose absence aids pathogenicity. The JGI analysis allowed a more stringent search for genes universally conserved in the pathogens and absent in the non-pathogenic strains since it allowed inclusion of all four genomes in the analysis.

One operon present in “*B. hampsonii*” strain 30446 and absent in “*B. hampsonii*” strain KL-180 comprised genes *nikA-E*, which are genes involved in nickel transport (Table 3.5). Nickel is an essential cofactor for a number of enzymatic reactions including production and consumption of hydrogen, hydrolysis of urea, oxidation of carbon monoxide, and detoxification of superoxide anion radicals (Eitinger and Mandrand-Berthelot, 2000). Nickel transport is thought to be important in the pathogenesis of *H. pylori*, where nickel is an essential cofactor of urease, thus acting as a virulence factor by enabling the hydrolysis of urea (Eitinger and Mandrand-Berthelot, 2000). It is possible the *nik* genes aid *Brachyspira* pathogenesis, and could be virulence factors.

Another operon found in strain 30446 but absent in KL-180 is the *sag* operon, which is responsible for the production of Streptolysin S (SLS). SLS is a cytolytic toxin and virulence factor in *Streptococcus* spp. (Molloy et al., 2011). This set of genes was of obvious interest because the *sag* operon is a well-established virulence factor of *Streptococcus* spp. where it acts as a hemolysin, breaking down erythrocytes of its host (Molloy et al., 2011), and haemolysis is associated with virulent *Brachyspira*. Pathogenic species *B. hyodysenteriae* and *B. hampsonii* are strongly haemolytic when cultured on blood agar, whereas non-pathogenic species are generally characterized by weak haemolysis, therefore it is likely haemolysis plays a role in virulence. In *Streptococcus*, the operon comprises a contiguous set of nine genes called *sagA* through *sagI*.

sagA encodes the SLS precursor peptide, followed by *sagBCD* which form an enzyme complex involved in post-translational modification (Lee et al., 2008). The enzyme complex contains a dehydrogenase, cyclodehydratase and a docking protein, and converts the *sagA* precursor molecule into SLS. The functions of the remaining proteins are not as well understood. *sagE* is thought to be a membrane spanning peptidase, and may be responsible for leader cleavage. *sagE* also has weak sequence similarity to a bacteriocin immunity protein found in *Lactobacillus plantarum* (Datta et al., 2005). The function of *sagF* has yet to be elucidated. *sagG*, *sagH*, and *sagI* are thought to form an ABC-transporter similar to those involved in bacteriocin export (Datta et al., 2005; Molloy et al., 2011; Nizet et al., 2000). Genes from the *sag* operon, except *sagA* and *sagF*, were found in the annotations from both RAST and JGI MGAP.

sagA forms the precursor protein which is modified to form the SLS molecule. Therefore, it is essential to the functioning of the operon, and without *sagA* SLS would not be a viable virulence factor in *Brachyspira* pathogenesis. This precursor protein is a small peptide, with a size of approximately 40 amino acids. This may be why the *sagA*-like gene was not identified in *Brachyspira* spp. through the annotated genomes. Issues have been identified with annotation of small proteins (less than 100 aa), where they can be missed resulting in an incomplete picture of an organism's capabilities (Hemm et al., 2008; Warren et al., 2010). The genome annotation in RAST identified hypothetical proteins in "*B. hampsonii*" strain 30446 as small as 114 bp, but the 123 bp *sagA* gene was not annotated.

PCR primers were designed for the *sag*-like genes, as a means to screen other *Brachyspira* species for the presence of these genes. It would be expected that the entire set of *sag* genes would

be universally present in pathogenic species of *Brachyspira* if it was an essential virulence factor. Epidemiologically distinct strains of “*B. hampsonii*”, that were isolated from samples derived from clinical cases of *Brachyspira*-associated disease, were screened by PCR for the *sag*-like genes. In seven of the strains, all the *sag*-like genes were present. *sagA* was not detected in D12-32529 and D13-06010, and in D13-06010 *sagG* and *sagI* were not detected. For these strains, since they were isolated from clinical cases, their pathogenicity has been assumed though has not been demonstrated as so in experimental inoculation trials. It is possible that strains D12-32529 and D13-06010 are non-pathogenic. Alternatively, the primers designed may not be completely universal since they were designed based only on “*B. hampsonii*” *sag* gene sequences. Similarly, type strains of *B. pilosicoli*, *B. murdochii*, *B. innocens* and *B. hyodysenteriae* were screened for the *sag*-like genes. Only in *B. pilosicoli* the *sagA* and *sagD* genes were not detected by PCR. All the genes were detected in the published whole genome sequence of *B. pilosicoli* P43/6/78 when a BLASTn search was performed, which suggests that the distribution of *sag* genes in *B. pilosicoli* is variable.

The detection of the *sag* operon in type strains of *B. murdochii*, *B. intermedia* and *B. innocens* was surprising, given that these species are weakly haemolytic and generally considered to be non-virulent (Jensen et al., 2010; Kinyon and Harris, 1979; Osorio et al., 2013; Phillips et al., 2010; Weissenböck et al., 2005). A limitation of this study was that only detection of ORFs, or gene presence, was determined but gene expression was not explored. Non-pathogenic *Brachyspira* strains may contain the *sag* genes but the genes may not be expressed, or expressed at low levels. Additionally, virulence is likely multi-factorial which may explain the presence of the *sag* genes in non-pathogenic species. The *sag* operon may not contain the full power of

pathogenicity, but rather play a part of the role. For example, *B. pilosicoli* causes a much milder disease than that caused by *B. hyodysenteriae* or “*B. hampsonii*”, and may only cause a milder disease because it lacks additional pieces required for full pathogenicity.

3.6 Conclusions

The basic genomic features of the pathogenic “*B. hampsonii*” strain 30446 and the non-pathogenic “*B. hampsonii*” strain KL-180, are similar and have an average nucleotide identity of 97%. Differences were identified through comparative genomics, and identifying gene content differences where there are genes unique to either strain, and many hypothetical proteins. Any gene differences could be important to disease outcome, including hypothetical proteins which have yet to be characterized and could represent novel virulence determinants. One set of genes uniquely present in the pathogenic strains (*B. hyodysenteriae*, “*B. hampsonii*” strains 30446 and 30599) and absent in the non-pathogenic strain (“*B. hampsonii*” strain KL-180) as identified through genome annotation, resembles the *sag* operon which is responsible for the production of Streptolysin S. The genes of this operon were found bioinformatically and biologically through PCR, and were found in pathogenic “*B. hampsonii*” strains. The presence of this operon in pathogenic *Brachyspira* species, and absence in the non-pathogenic “*B. hampsonii*” strain KL-180 supports the possibility of the operon being a virulence factor and contributing to disease outcome.

3.7 Transition Statement

The characterization of pathogenic and non-pathogenic strains of “*B. hampsonii*” demonstrated that putative virulence factors exist in the pathogenic strain, and these virulence factors could contribute to the progression of disease. However, pigs experimentally inoculated with pathogenic strains of *Brachyspira* do not all develop clinical disease and pathogenic strains have also been isolated from pigs in the field that are clinically healthy. The host, microbial and environmental factors that influence disease progression in exposed animals are not known. Our current diagnostic methods regularly detect multiple *Brachyspira* species in clinical samples from affected pigs. It is possible that mixed *Brachyspira* infections, or multi-species *Brachyspira* “communities”, may contribute to disease development. Investigation of this possibility requires an initial determination of the frequency of occurrence of mixed infections in clinical cases using an objective approach that allows simultaneous detection of species in the fecal microbiome that may be present across a wide range of abundance levels. Exhaustive detection of *Brachyspira* species in clinical samples from a range of production systems in different countries will also provide insight into the global distribution of “*Brachyspira hampsonii*”.

4 Characterization of *Brachyspira* communities from clinical cases of swine mucohaemorrhagic diarrhea

4.1 Abstract

The re-emergence of *Brachyspira*-associated disease and novel causative species “*B. hampsonii*” is a concern for swine producers. Current diagnostic methods have limitations and often utilize multiple techniques to provide a more accurate description, sometimes resulting in the detection of multiple *Brachyspira* spp. in samples from individual pigs. The pathogenesis of *Brachyspira*-associated disease is not completely understood, and it is not known whether multiple *Brachyspira* spp. are important in disease development. Deep sequencing of the *nox* gene target was used to simultaneously detect *Brachyspira* spp. in fecal samples from clinical cases of mucohaemorrhagic diarrhea in pigs from Western Canada, Mexico and Brazil. Synthetic mixtures of *Brachyspira* genomic DNA were created and sequenced to aid interpretation of clinical case sequence data and define detection limits for the technique. Samples from Brazil were dominated by *B. hyodysenteriae* and samples from Mexico were dominated by *B. hyodysenteriae* and *B. pilosicoli*. “*B. hampsonii*” strains were detected in these samples, though comprised a small proportion of the communities and are likely not clinically significant or the driving cause of disease. Samples from Canadian clinical cases were dominated by different species, including “*B. hampsonii*” strains. Sequencing often detected multiple species in individual samples, indicating that multiple species *Brachyspira* infections are common. Deep sequencing of the *nox* gene target is a suitable method for simultaneous detection of *Brachyspira* from clinical case material that may offer advantages over current, more targeted approaches.

4.2 Background

Brachyspira is associated with swine dysentery, or mucohaemorrhagic diarrhea and colitis in pigs. The disease, which was traditionally associated with *Brachyspira hyodysenteriae* (Harris et al., 1972), had been mostly eliminated from Canadian swine production systems by the early 1990s. Of growing concern is the re-emergence of disease in the late 2000s that has been associated with a novel species, “*Brachyspira hampsonii*” (Chander et al., 2012; Harding et al., 2010). While there is renewed interest in developing methods to treat and control swine dysentery, there are major challenges in mitigating this production-limiting disease including a lack of robust clinical diagnostic methods and limited understanding of the complex ecology of *Brachyspira* species.

Diagnostic methods currently include selective culture for *Brachyspira* spp., gross pathology, histopathology, serology, microscopic detection, biochemical assays of isolates, and PCR (Burrough, 2016). Limitations exist with different methods, for example the success of selective culture depends on the nature of the clinical sample and proper transportation and storage. PCR methods may lack analytical sensitivity or may not detect novel pathogenic species. Due to limitations with various methods, diagnostic protocols may employ multiple methods to detect and/or identify the presence of *Brachyspira* spp.. We and others have observed that application of multiple diagnostic methods often results in detection of multiple species from individual clinical samples. A study describing the prevalence of weakly haemolytic *Brachyspira* in pigs found that concurrent infections with two or more *Brachyspira* spp. was common, and weakly haemolytic species may contribute to colonic pathology (Komarek et al., 2009). It is not known whether mixed infections are of clinical significance in regards to swine dysentery, and whether multiple species may contribute to the pathogenesis of the disease.

Due to the lack of robust diagnostic methods available for *Brachyspira* detection and gaps in our knowledge about mixed species infections, this study aimed to develop a method for simultaneous detection of *Brachyspira* species in clinical samples using deep sequencing of PCR amplicons from the NADH oxidase (*nox*) gene. This method was then used to describe mixed infections from pigs with swine dysentery, and aid in our understanding of the distribution of *Brachyspira* spp..

4.3 Methods

4.3.1 *Brachyspira* isolates, culture and DNA extraction

The *Brachyspira* strains used in this study included *B. hyodysenteriae* ATCC 27164^T, *B. pilosicoli* ATCC 51139^T, “*B. hampsonii*” strain 30446, and “*B. hampsonii*” strain 30599. The type strains of *B. hyodysenteriae* and *B. pilosicoli* were obtained from the American Type Culture Collection. “*B. hampsonii*” 30446 and 30599 were originally isolated from diarrheic pigs in Western Canada (Costa et al., 2014b; Rubin et al., 2013a). These *Brachyspira* isolates have been stored long-term at -80°C in brain-heart infusion (BHI) broth (BHI, Becton Dickinson Canada, Mississauga, ON) containing 10% (v/v) glycerol. Strains from storage were cultured in JBS broth (BHI with 1% (w/v) glucose, 5% (v/v) fetal bovine serum, 5% (v/v) sheep blood), and the cultures were incubated anaerobically at 42°C with stirring.

Genomic DNA was extracted from cultures of *Brachyspira* by pelleting 1 mL of culture and following a modified salting-out procedure, as previously described (Martin-Platero et al.,

2007). Concentration of the extracted DNA was determined using a Nanodrop spectrophotometer, and the DNA was stored at -20°C.

4.3.2 Clinical samples

Canadian samples were from cases of mucohaemorrhagic diarrhea that were submitted to our laboratory for diagnostics. Since 2009, clinical cases have been submitted from farms in Western Canada to determine if *Brachyspira* is present, and if so, what species, through culture, PCR and sequencing methods. This has resulted in a collection of tissues and/or isolates from nearly 1000 clinical cases, and from many of these cases *nox* sequences were obtained and are available in a sequence database maintained by our research group. From these cases, we chose samples to include in this study based on whether they were *Brachyspira* positive by any diagnostic method, and epidemiologically distinct, or from different farms. We included cases with results that suggested both infections caused by a single *Brachyspira* species, as well as possible mixed species infections. Upon original submission of fecal samples, total genomic DNA was extracted using the QIAamp DNA Stool Mini Kit, and then stored at -20°C. Archived fecal DNA extracts were available for this study.

Clinical case material from Brazil was generously donated by Roberto Guedes (Federal University of Minas Gerais). Fecal DNA samples collected from 20 pigs with mucohaemorrhagic diarrhea were received. Samples were received as DNA extracts in solution, and were shipped on dry ice. DNA was extracted in Brazil using the QIAamp DNA Stool Mini Kit, using 200 mg of feces per reaction.

Clinical case material from Mexico was generously donated by Enrique Corona-Barrera (Universidad de Guanajuato). Fecal DNA extracts (n = 46) and DNA extracts from cultured isolates (n = 11) were received. All were collected from pigs with mucohaemorrhagic diarrhea and that were culture positive for *Brachyspira*. Each sample was from a different pig, and samples were from 39 different farms. The DNA was extracted in Mexico using the Qiagen QIAamp DNA Stool Mini Kit and stored at -80°C. DNA extracts were dehydrated using a SpeedVac for 30 minutes, and dehydrated DNA samples were shipped. Once received, DNA was rehydrated with 50 µl of TBE buffer, and stored at -20°C.

The suitability of genomic DNA extracts from clinical samples for PCR was confirmed using PCR targeting the pig cytochrome oxidase subunit 1 (*cox1*) gene. The primers are listed in Table 4.1, and each PCR reaction contained 1 x PCR reaction buffer (0.2 M Tris-HCl at pH 8.4, 0.5 M KCl), 2.5 mM MgCl₂, 200 µM dNTP, 400 nM each primer, 2.5 U Taq DNA Polymerase and 2µl of template DNA, in a final volume of 50 µl. PCR was performed in a thermocycler (Eppendorf Mastercycler) with the following program: 94°C for 5 min, 40 cycles of 95°C for 30 s, annealing at 60°C for 30 s, and final extension of 72°C for 2 min. A no template control and positive amplification control (pig fecal DNA extract from an unrelated study) were included. PCR products were visualized on a 1% agarose gel by ethidium bromide staining. Only samples producing the expected PCR product were included in the study.

4.3.3 Construction of synthetic *Brachyspira* communities and quantitative real-time PCR

Synthetic microbial communities were created by spiking known amounts of *Brachyspira* DNA into a background of fecal DNA known to be *Brachyspira*-free. To determine if the background fecal DNA contained *Brachyspira*, fecal DNA extracts (obtained from an unrelated study) were screened with PCR targeting the *nox* gene as previously described (Rohde et al., 2002). All negative samples were pooled together to create a uniform fecal DNA background.

Quantification of genomic DNA from four *Brachyspira* isolates (*B. hyodysenteriae* ATCC 27164^T, *B. pilosicoli* ATCC 51139^T, “*B. hampsonii*” strain 30446, and “*B. hampsonii*” strain 30599), was achieved by qPCR and allowed known amounts of genomic DNA from each species to be added to the synthetic communities. Quantitative real-time PCR (qPCR) was performed using the primers described in Table 4.1, and as previously described. Briefly, qPCR was performed using the Bio-Rad MyiQ thermocycler with iQ SYBR green supermix (Bio-Rad Laboratories (Canada) Ltd., Mississauga, Ontario). All samples were tested in duplicate, and no template controls were included with each run. Quantification was achieved by using a serial dilution of a standard curve of plasmids containing the target sequence. After quantification, samples were diluted to produce solutions containing high, medium, or low levels of genomic DNA from each *Brachyspira* species. Two microliters of each dilution were added to the fecal DNA background (100µl fecal DNA) to produce *Brachyspira* “communities”. Three types of communities were created: communities with high levels (10^5 copies/µl) of one *Brachyspira* species and low levels (10^1 copies/µl) of the other three, communities with high levels of one species and medium levels (10^3 copies/µl) of the other three, and communities with equal levels (high, medium or low) of

each of the four species. Synthetic communities with a high level of one species with medium or low levels of the remaining three species were created to determine if we can detect mixed infections in a fecal background with variable levels of different species. Communities with all four species at high, medium, or low levels were created to elucidate whether there is any PCR bias with the *nox* primers. The synthetic communities underwent qPCR, as described above, to verify their composition.

Table 4.1. Primer sequences used in this study.

Gene Target	Application	Primer Name	Primer Sequence (5'-3')	Annealing temperature	Product size	Reference
<i>cox1</i>	Pig specific	JH0183	TCA TCG GGG CCC TAC GAG CA	60	278	(Hill, unpublished)
		JH0184	GGC GAA AGG TCC GGC TGC AT			
<i>cpn60</i>	<i>Brachyspira hyodysenteriae</i> SYBR assay	JH0073	AGT GAA ATA GTT GCT CAT ATC AAA T	63	120	(Rubin et al., 2013a)
		JH0074	GCA TCA CTG ATT AAA GAA CCA AT			
<i>cpn60</i>	<i>Brachyspira pilosicoli</i> SYBR assay	JH0077	ACA ATG ATA AAG AGA TAG GTG CTT	60	111	(Rubin et al., 2013a)
		JH0078	CTA ATG AAA GGC TAG TTT CTA ATG AT			
<i>nox</i>	" <i>Brachyspira hampsonii</i> " 30446 SYBR assay	JH0224	TCG CTA AAT TAT TCC AAC AAG GA	60	215	(Rubin et al., 2013a)
		JH0225	AAA CGC ATT TCT ATT CCA GCA			
hypothetical protein (Genbank accession WP_008726773)	" <i>Brachyspira hampsonii</i> " 30599 SYBR assay	JH0436	AAA GTG CCA CAG GCA ATG TA	63	176	(Costa et al., 2014b)
		JH0437	TGC AAG ATT AGA CGG AGC AA			
<i>nox</i>	<i>Brachyspira</i> genus specific (with Illumina adapters added for deep sequencing)	noxF_adaptor	TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG TAG CYT GCG GTA TYG CWC TTT GG	59	939	(Rohde et al., 2002)
		noxR_adaptor	GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GCT TCA GAC CAY CCA GTA GAA GCC			

4.3.4 Deep Sequencing and analysis

Amplicons for deep sequencing of clinical samples and synthetic communities were produced through PCR using primers that are *Brachyspira* genus-specific, and target the *nox* gene (Rohde et al., 2002). PCR was performed as previously described, except Illumina adapters were added to the original primers (Table 4.1).

The library preparation for the amplicons was performed following the 16S Metagenomic Sequencing Library Preparation protocol (Illumina, San Diego, CA). Amplicon PCR product clean-up was performed using the Qiaex II gel extraction kit from products visualized on a 1% agarose gel by ethidium bromide staining. The Index PCR product clean-up was performed as described in the protocol except 36 μ l of the Ampure beads were used, due to the length of the amplicon being sequenced (1052 bp with 939 bp from the *nox* product and 113 from Illumina-adapted primers). Average size of the final amplicon libraries was verified using the DNA High-Sensitivity chip on the 2100 Bioanalyzer (Agilent Technologies, Inc., Santa Clara, CA) and the concentration was measured using the Qubit (Life Technologies, Inc., Burlington, ON). Sequencing was performed on the Illumina MiSeq using the 500 cycle (2 x 250) V2 Nano sequencing kit. PhiX DNA was loaded as an internal control as described, except 20% (v/v) was added to the library due to the expectation of low diversity libraries.

Trimmomatic was used to process raw sequence reads using the following criteria: minimum length of 50 bp, average quality score of 15, and leading and trailing low quality of 3

(Bolger et al., 2014). Quality trimmed reads were mapped on to a set of seven reference sequences using Bowtie2 (Langmead and Salzberg, 2012). Reference genome sequences were obtained from NCBI Genbank for these seven *Brachyspira* species: *B. hyodysenteriae* strain WA1, “*B. hampsonii*” strain 30446, “*B. hampsonii*” strain 30599, *B. intermedia* strain PWS/A, *B. pilosicoli* strain P43/6/78, *B. murdochii* strain DSM 12563, and *B. innocens* strain B256. From the genomes, the full sequence of the *nox* gene were extracted for inclusion in the Bowtie2 index. Resulting SAM files were parsed to determine the abundance of reads corresponding to each species in each sample.

Jackknifed Bray-Curtis dissimilarity was calculated using the Quantitative Insights Into Microbial Ecology (QIIME) package (Caporaso et al., 2010) to assess beta diversity. A Principal Coordinate Analysis plot (PCoA) was generated using the to visualize the Jackknifed Bray-Curtis dissimilarity, where each point in the PCoA plot represents a clinical case. Samples that contain the most similar *Brachyspira* communities cluster together. Comparison of the *Brachyspira* community composition between samples was clustered by Jensen-Shannon, and heatmap3 package of R to generate the heatmap image (Zhao et al., 2014).

4.4 Results

4.4.1 *nox* amplicon sequence similarity

The genus-specific primers targeting the *nox* gene produce an amplicon with a size of 939 bp, and with the added Illumina-adapted primers 1052 bp. When sequencing amplicons with the Illumina MiSeq platform, the technology produces relatively short read lengths, and so is not

conducive to the 1052 bp amplicon. Sequences would be obtained on each end of the amplicon, though it is not known how informative each end would be or how suitable the ends would be for species resolution compared to the full amplicon sequence. Bioinformatically, we compared the sequence similarity of the full-length *nox* amplicon, the first 200 bp and the last 200 bp.

All *nox* sequences were obtained from our database of clinical and reference strains (n=212), and were aligned using ClustalW (Larkin et al., 2007). The aligned sequences were trimmed to 869 bp, and any shorter sequences were excluded leaving 155 sequences after trimming. BLASTClust was used to remove identical sequences, resulting in 70 unique *nox* sequences. Pairwise similarity, or percent identity, between the 70 unique *nox* sequences was determined using GeneDoc. Pairwise comparisons were done for the full amplicon sequences, the first 200 base pairs and the last 200 base pairs. The distributions of pairwise sequence identities are shown in Figure 4.1.

The full *nox* amplicon had a median pairwise percent identity of 90%, an average percent identity of 90% and a standard deviation of 3.2%. The first 200 bp of the *nox* amplicon had a lower median percent identity at 88% (average of 89% with a standard deviation of 4.3%), while the last 200 bp had a higher median percent identity at 93% (average of 92% with a standard deviation of 4.7%). The full *nox* sequences were selected to be unique, and therefore there were no identical pairwise comparisons of 100% in this category. The first and last 200 bp of the *nox* sequence had 45/2415 and 38/2415 comparisons with 100% identity, respectively. These results indicate that in 98% of cases, 200 bp from either end of the *nox* amplicon sequence is sufficient to discriminate among these 70 *Brachyspira* isolates.

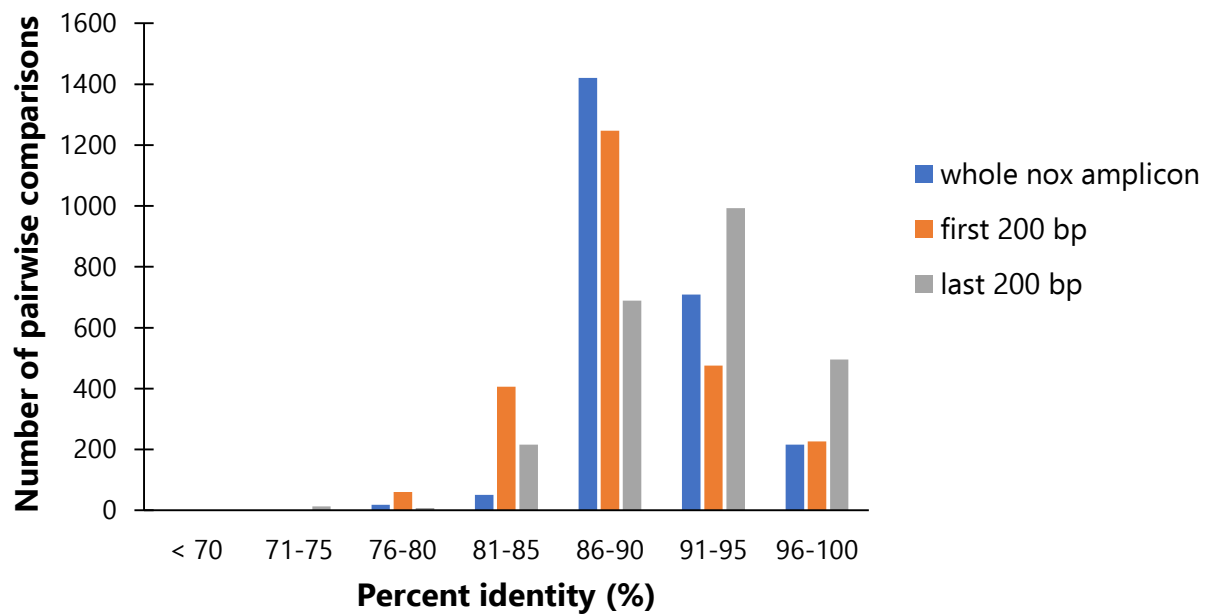


Figure 4.1. The percent identity of 70 pairwise comparisons of full-length *nox* gene sequences, the first 200 bp, and the last 200 bp.

4.4.2 Limit of detection of conventional *nox* PCR and synthetic communities

Conventional *nox* PCR is commonly used to detect and identify *Brachyspira* within clinical cases of mucohaemorrhagic diarrhea. To utilize deep sequencing of the *nox* target for diagnostic purposes, it is required to be able to produce visible amounts of PCR product for purification and sequencing. Understanding the lower limits of *Brachyspira* required to be detectable by conventional *nox* will also aid in interpreting what detection through deep sequencing means in terms of actual abundance.

To determine the visual detection limits of conventional *nox* PCR, various concentrations (10^0 to 10^7 copies/ μ l) of *Brachyspira* genomic DNA were used in the standard *nox* PCR protocol, and the products were visualized on an agarose gel. Four different *Brachyspira* species were used: *B. hyodysenteriae*, *B. pilosicoli*, “*B. hampsonii*” strain 30446 and “*B. hampsonii*” strain 30599. To determine the effects of a fecal DNA background on detection limits, *Brachyspira* genomic DNA dilutions of each species were either spiked with a fecal DNA background from a *Brachyspira*-negative pig or not, and *nox* PCR was performed and products were visualized on an agarose gel (Figure 4.2). Lower levels of *Brachyspira* DNA could be detected without the fecal DNA background, than with the fecal DNA background. The lowest level detected 10^2 copies/ μ l of *B. pilosicoli* with no fecal DNA background.

The synthetic communities were created with high, medium, or low levels of four different *Brachyspira* species. High levels of *Brachyspira* contained 10^5 copies/ μ l, medium levels 10^3 copies/ μ l and low 10^1 copies/ μ l. To verify the composition of the synthetic communities, species specific qPCR was performed, and the amount detected in each community reflected what was spiked in (Table 4.2).

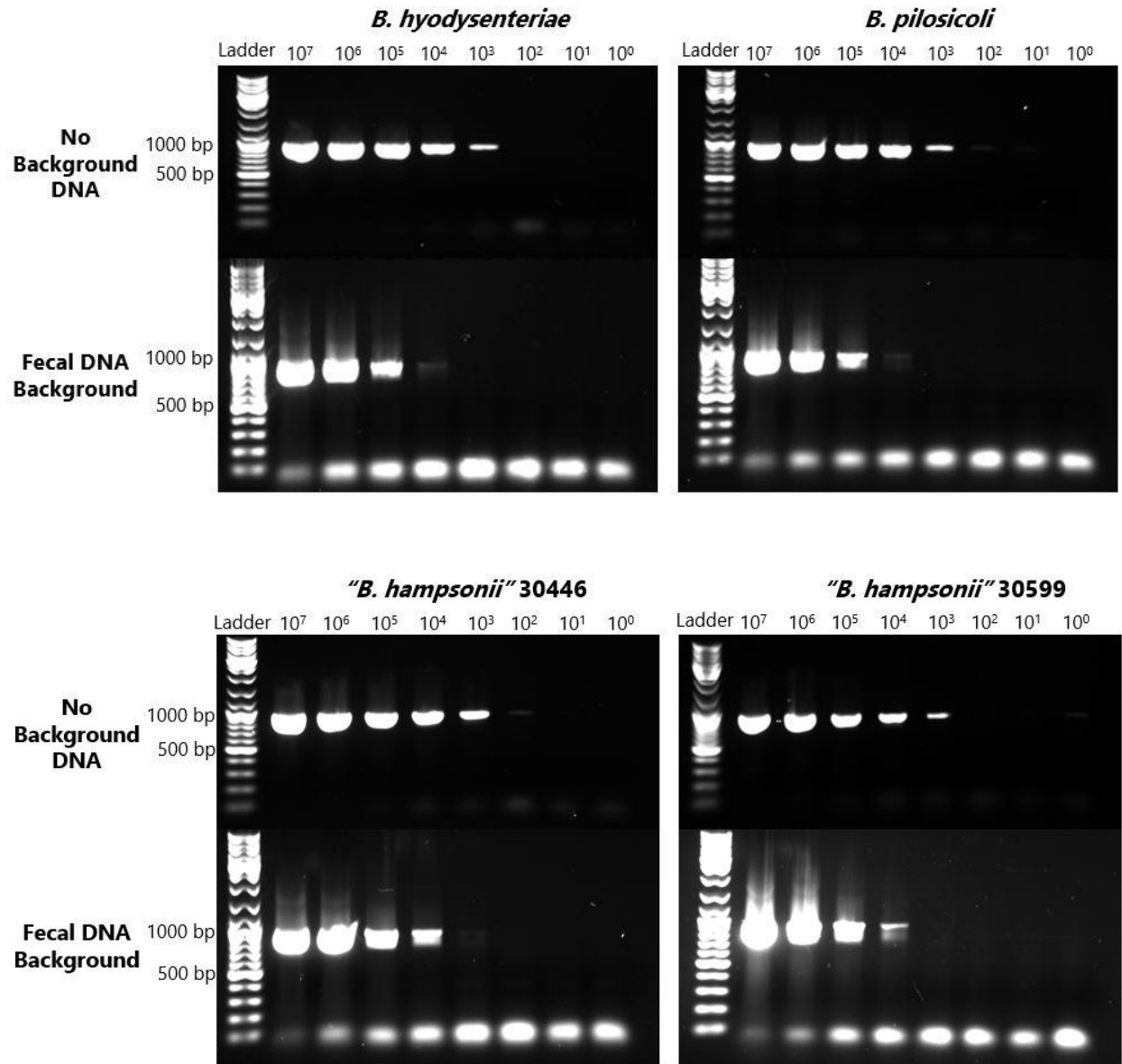


Figure 4.2. *B. hyodysenteriae*, *B. pilosicoli*, "*B. hampsonii*" 30446, and "*B. hampsonii*" 30599 genomic DNA dilution curves from 10^7 to 10^0 copies/ μ l visualized on an agarose gel.

The standard *nox* PCR protocol was used on a dilution series of genomic DNA from four *Brachyspira* species, with no additional DNA and with a fecal DNA background spiked in.

Table 4.2. The amount of each *Brachyspira* species in the synthetic *Brachyspira* communities as determined by qPCR.

The rows highlighted with green indicate high levels of one *Brachyspira* species, and low levels of the other three species were added to the synthetic community. Purple indicates high levels of one species, and medium of the remaining three. Blue indicates all four species were added at the same level. In S9 all four species were added at high levels, S10 at medium levels, and S11 at low levels. The bold numbers indicate high levels of that *Brachyspira* species were added in the identified synthetic community.

Synthetic Community	Log ₁₀ copies* per µl			
	<i>B. hyodysenteriae</i>	<i>B. pilosicoli</i>	<i>"B. hampsonii"</i> 30446	<i>"B. hampsonii"</i> 30599
S1	6.7	2.4	1.8	2.1
S2	2.4	6.0	2.0	1.7
S3	3.4	1.9	6.1	1.7
S4	2.3	1.8	1.9	5.5
S5	6.9	3.9	4.0	3.4
S6	4.6	6.0	4.1	3.2
S7	4.7	3.9	6.1	3.2
S8	4.6	4.0	4.0	5.2
S9	6.8	6.2	6.2	5.3
S10	4.7	4.0	4.1	3.2
S11	2.7	1.8	1.8	1.8

*Average of two technical replicates.

4.4.3 Deep sequencing of synthetic communities and clinical samples

Deep sequencing of both the synthetic communities and the clinical cases was performed on the Illumina MiSeq. The method was used to sequence *nox* amplicons from the fecal DNA extracts of 110 clinical cases of mucohaemorrhagic diarrhea and 11 synthetic communities containing different ratios of four *Brachyspira* species as described in Table 4.2. The samples were divided into two groups, each containing 55 samples for two separate sequencing runs. Each sequencing run contained the 11 synthetic communities, resulting in each run containing a total of 66 samples and the synthetic communities being sequenced twice.

For the first run, a total of 426,943 sequence reads were obtained, ranging from a minimum of 130 reads to a maximum of 81,647 reads, and an average of 6469 reads per sample. After quality trimming, a total of 405,069 reads remained for analysis. The second run produced a total of 582,742 sequence reads, ranging from a minimum of 3009 reads to a maximum of 26,371 reads per sample and an average of 8829 sequence reads. After quality trimming, 512,597 sequence reads remained for analysis. For each run, only samples with a minimum of 1000 reads were used in further analysis.

Mapping the sequence reads using Bowtie2 on to the set of seven reference *nox* sequences resulted in 4% to 99% of reads mapped per sample. Manual inspection of a subset of unmapped reads indicated that they correspond to non-target sequences. Samples with mapping rates below 55% were excluded from further analysis. After mapping, sequences from 89 clinical samples and the 11 synthetic communities (sequenced in duplicate) remained for further analysis.

4.4.4 *Brachyspira* communities from synthetic communities

From the Bowtie2 results, we obtained species identification of *Brachyspira* present and their abundance from each sample (Figure 4.3). The first four synthetic communities (S1 to S4) contained one *Brachyspira* species spiked in at a high level, and the additional three species at low levels. The *Brachyspira* community determined by deep sequencing reflected its composition, where the species spiked in at high levels comprised at least 98% of the community in each case. Synthetic communities S5 to S8 contained one species spiked in at high levels, and the remaining species at medium levels. Similarly to communities S1 to S4, communities S5 to S8 contained the species spiked in at high levels as the highest proportion of the *Brachyspira* community, except in the case of “*B. hampsonii*” strain 30599, where it was not the dominant species. Communities S9, S10 and S11 contained all four *Brachyspira* species at high, medium, and low levels, respectively. In these communities, all four species were detected though there was negative bias towards “*B. hampsonii*” strain 30599. In the three communities, “*B. hampsonii*” strain 30599 comprised 1-2% of the community, despite all four species were present at comparable levels.

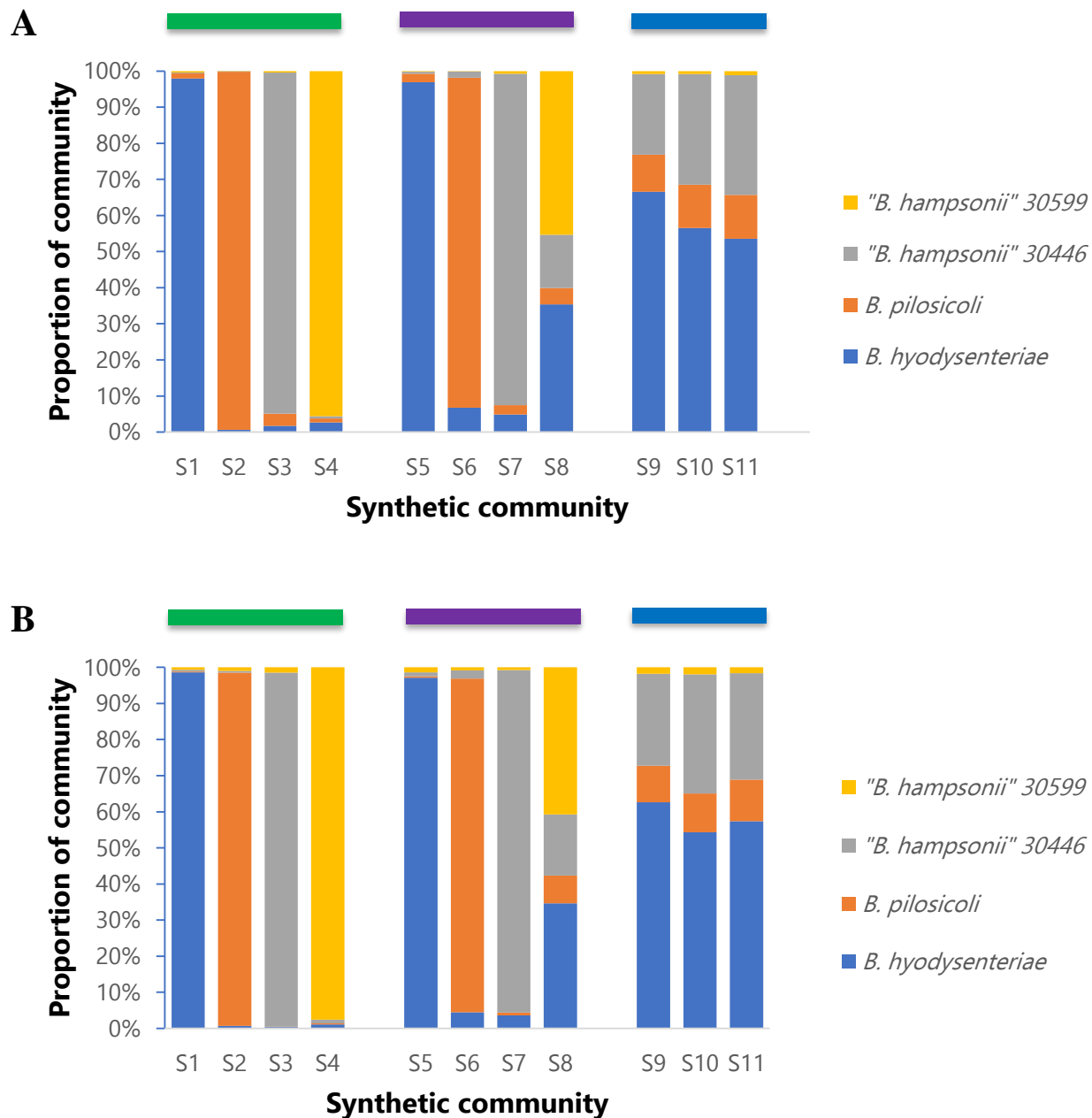


Figure 4.3. Proportion of four different *Brachyspira* species present in synthetically created communities.

Two separate sequencing runs shown as A and B. The colored bars above the columns indicate the structure of the synthetic community. Bars with green indicate high levels of one *Brachyspira* species, and low levels of the other three species were added to the synthetic community. Purple indicates high levels of one species, and medium of the remaining three. Blue indicates all four species were added at the same level. In S9 all four species were added at high levels, S10 at

medium levels, and S11 at low levels. The bold numbers indicate high levels of that *Brachyspira* species were added in the identified synthetic community.

4.4.5 *Brachyspira* communities from clinical cases

Mexican samples were dominated by *B. hyodysenteriae* or *B. pilosicoli* (Figure 4.4). The samples from Brazil were all dominated by *B. hyodysenteriae* and had proportionally very little of any other *Brachyspira* species (Figure 4.5). Over 98% of all the *Brachyspira* detected in Brazilian samples was *B. hyodysenteriae*. In all the samples from both Mexico and Brazil, however, there were multiple species detected in every clinical sample, ranging from 4 to 7 different species, including “*B. hampsonii*”. The maximum proportion of “*B. hampsonii*” detected, however, was only 0.7% in the Brazilian samples. Somewhat more appreciable levels were detected in the Mexican samples, ranging from 0.1-29% in all 31 samples where it was detected.

Cultured isolates from four of the Mexican clinical cases (M21, M22, M30 and M31) had been provisionally identified as *Brachyspira* spp.. Standard *nox* PCR was performed on provided DNA extracts from these isolates using genus-specific primers as previously described (Rohde et al., 2002), to verify the culture samples were *Brachyspira*. PCR products were purified using the Qiagen PCR product purification kit and sequenced using the amplification primers, and sequences were compared using BLASTn to our *Brachyspira nox* sequence database for species identification. The four culture isolates were all identified as *B. hyodysenteriae*. *B. hyodysenteriae* comprised 79%, 54%, 47% and 29% of the total *Brachyspira* detected by deep sequencing of *nox* amplicons from the corresponding clinical samples, respectively (Figure 4.4).

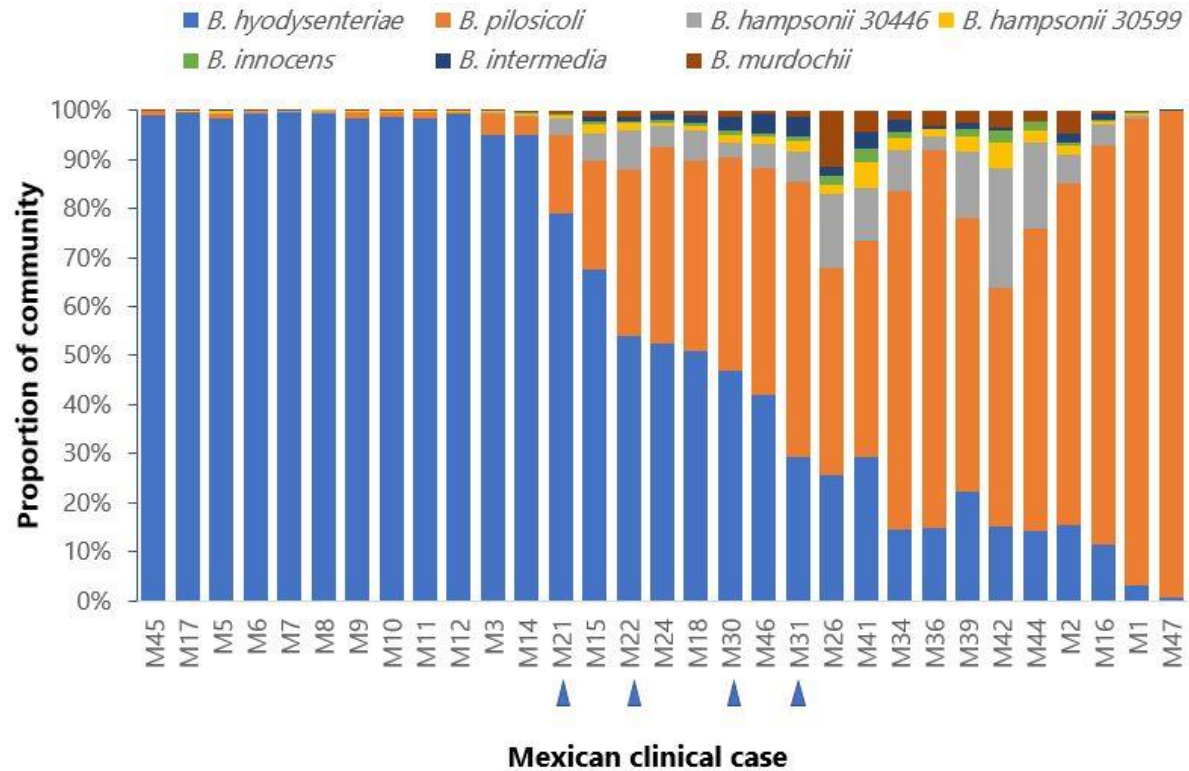


Figure 4.4. Proportion of each community that is comprised of different *Brachyspira* species for each Mexican clinical case (n=31).

The coloured triangles represent the *Brachyspira* species identified through sequencing of the *nox* gene for culture isolates. For example, the case M20 had an accompanying culture isolate that was identified as *B. hyodysenteriae*.

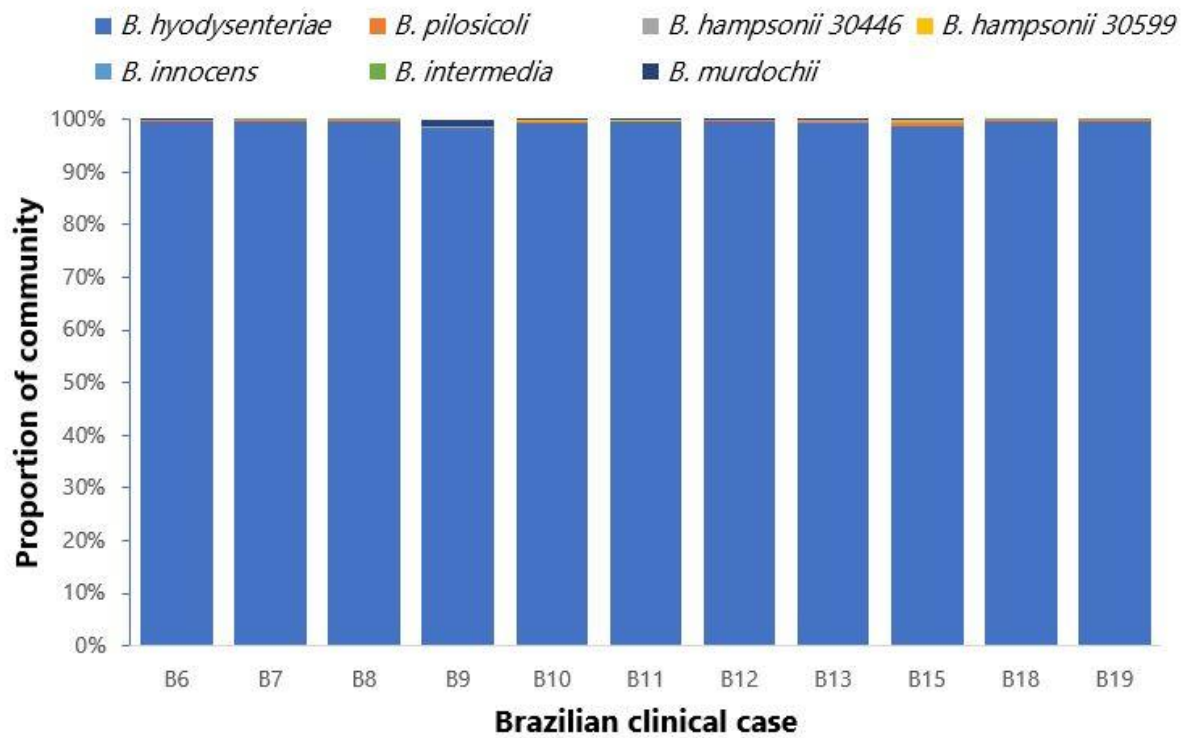


Figure 4.5. Proportion of each community that is comprised of different *Brachyspira* species for each Brazilian case (n=11).

The Canadian samples were dominated by a range of different *Brachyspira* species (Figure 4.6). “*B. hampsonii*” strains dominated, or composed over 50% of the community, in 18 samples. For the Canadian cases, we had access to records for the original diagnosis of which species were detected when the sample was originally submitted. The results from the database are represented as the coloured triangles below the bars in Figure 4.6. In all cases, the species detected in the original diagnostic testing were also detected by deep sequencing of *nox* amplicons. In 43 out of 47 cases, the species reported in the original diagnostic investigation accounted for over 50% of the sequence reads, or dominated the profile. Diagnostic testing of clinical cases C139 and C68 identified “*B. hampsonii*” strain 30599, but the community profiles from these samples were not dominated by this species, which accounted for 49% and 31% of the communities, respectively. *B. innocens* was found diagnostically in clinical case C114, and sequences corresponding to this species comprise 41% of the community profile for this sample. The community profile of clinical case C97 was dominated by “*B. hampsonii*” strain 30599 at 76%, however “*B. hampsonii*” strain 30599 was not identified during the original diagnostic investigation. Instead, diagnostics identified *B. pilosicoli* and *B. intermedia*, which both combined accounted for only 4% of the community as determined through deep sequencing.

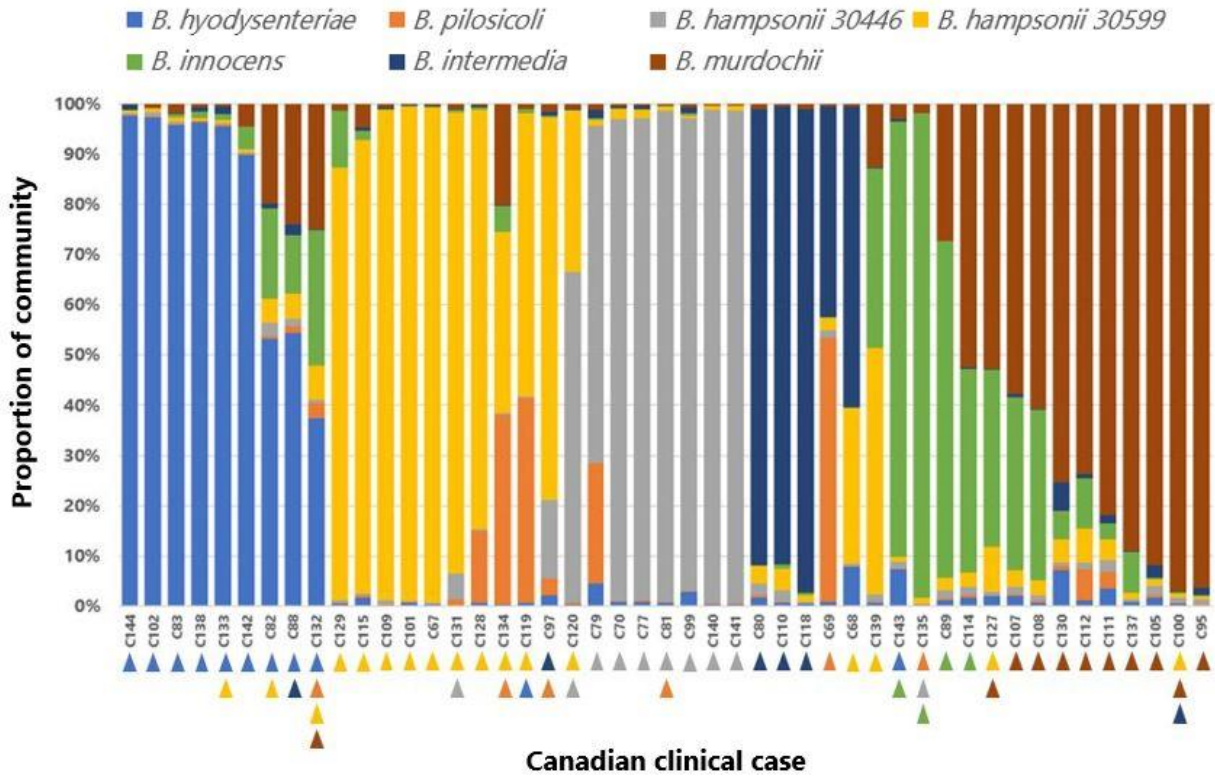


Figure 4.6. Proportion of each community that is comprised of different *Brachyspira* species for each Canadian clinical case (n=47).

The coloured triangles represent the reported diagnosis based on selective culture and PCR diagnostics. For example, the first case C144 was originally diagnosed as *B. hyodysenteriae*.

Cluster analysis of the clinical cases indicated they differ by country of origin (Figure 4.7 and 3.8). The heatmap demonstrates the communities cluster based on which *Brachyspira* dominates the community. *B. hyodysenteriae* dominated profiles were identified in samples from all three countries, while “*B. hampsonii*” dominated profiles were observed only in Canadian samples. The PCoA plot (Figure 4.8) further illustrates the greater diversity among profiles of the Canadian samples than either the Mexican or Brazilian samples.

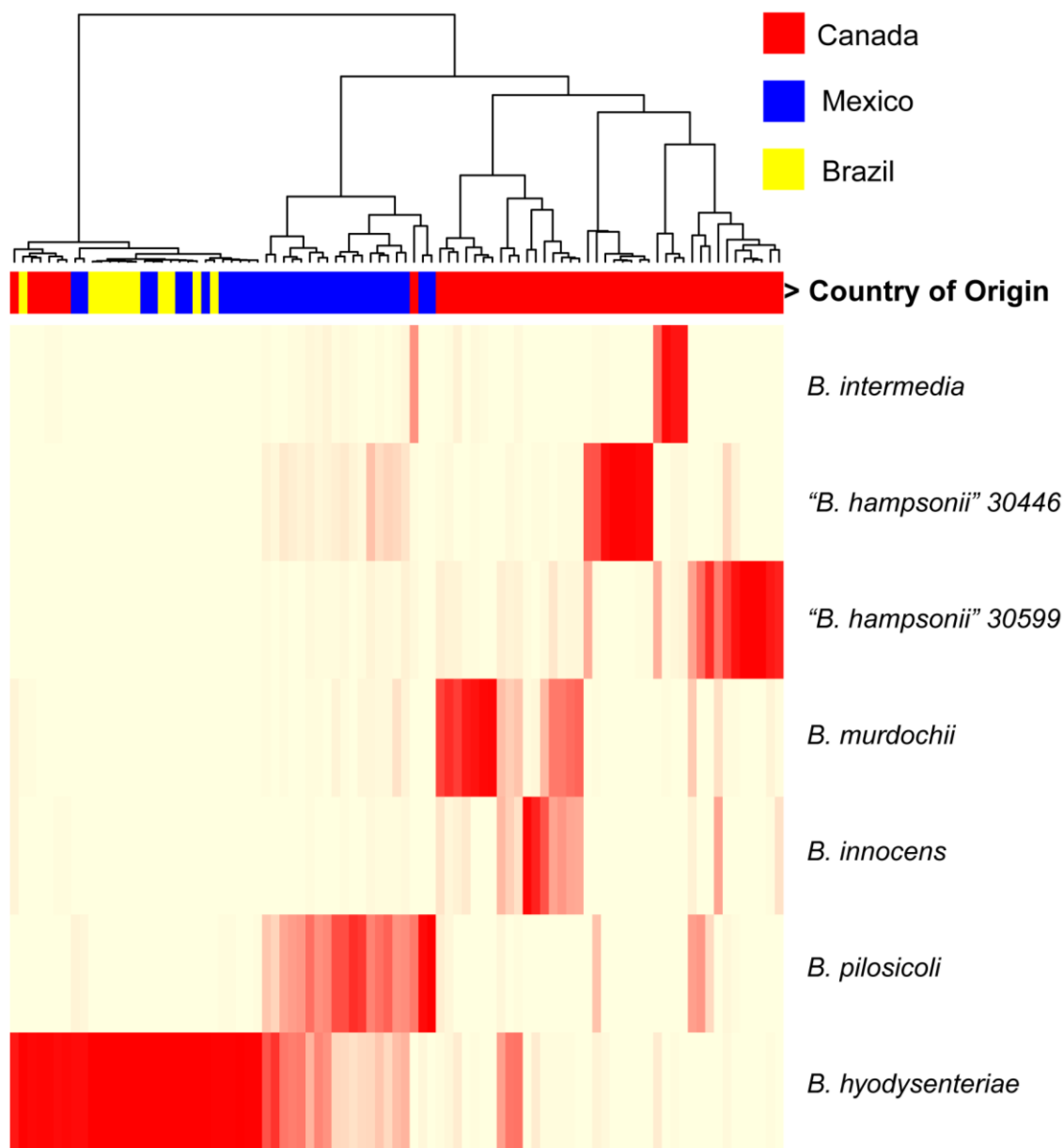


Figure 4.7. Cluster analysis of *Brachyspira* communities based on the proportion of different *Brachyspira* species present in each sample.

Each column, or branch of the tree, represents a sample and the colored bars for each column indicate which country the sample originated. The shading for each *Brachyspira* species demonstrates the relative proportion present in the community, where darker shades of red represent a high proportion.

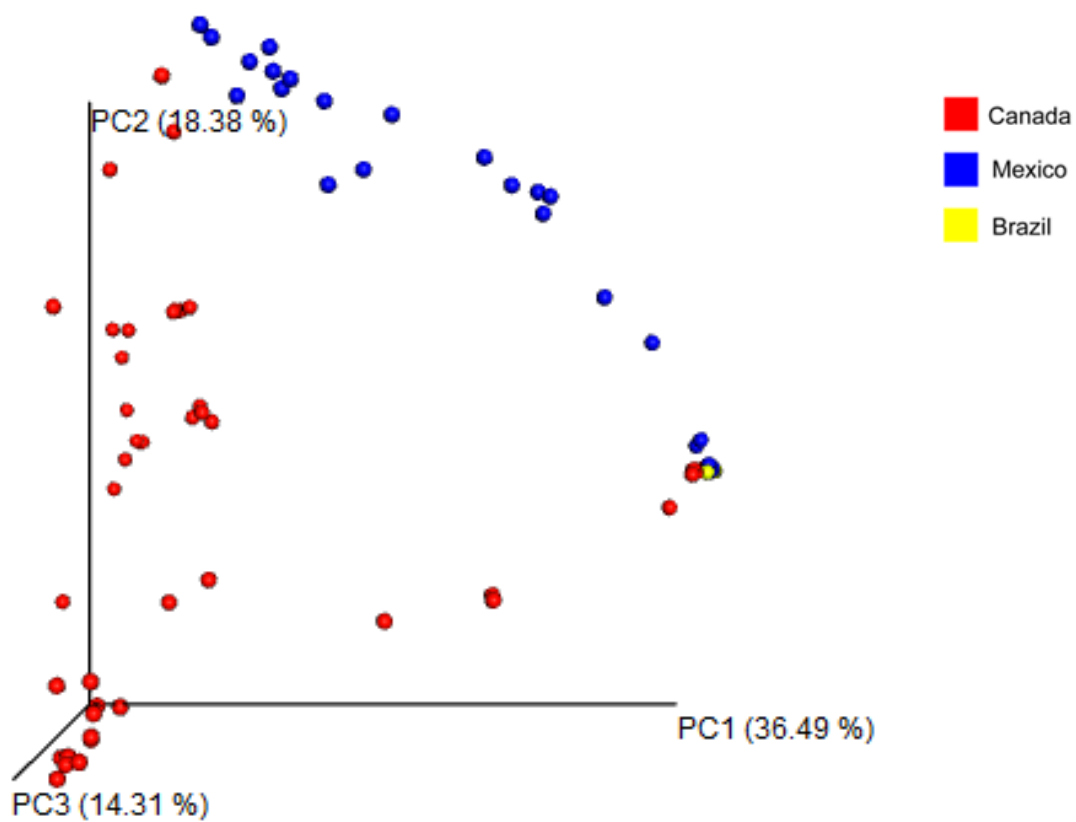


Figure 4.8. Principal Coordinates Analysis based on Bray-Curtis dissimilarity values. Variable include the country of origin of the clinical cases, and the distribution demonstrates sample similarity.

4.5 Discussion

Sequencing of microbial communities, using gene targets such as the *cpn60* universal target or the 16S rRNA gene, is a common way to get a view of the microbiome or a big-picture view of the types of bacteria present in different samples. This method targeting the *cpn60* universal target has previously been used to investigate the fecal microbiota of pigs with mucohaemorrhagic diarrhea resulting from experimental inoculation with “*B. hampsonii*”. It was found that *Brachyspira* is present at such low levels, with respect to rest of the microbial community, that the number of sequence reads in the data corresponding to *Brachyspira* was extremely low (Costa et al., 2014a). Thus, to investigate the types of *Brachyspira* present in the microbial community of affected pigs, a more selective approach was needed to enrich PCR product libraries for *Brachyspira* sequences. We followed the same sequencing method with a different gene target: the NADH-oxidase gene, or *nox*, which is commonly used for *Brachyspira* detection and identification.

The *nox* gene target, to our knowledge, has not been previously used for microbiome sequencing though it has been used extensively for *Brachyspira* species identification (Chander et al., 2012; Fellstrom et al., 2008; Hidalgo et al., 2010; Jansson et al., 2008; Martinez-Lobo et al., 2013; Patterson et al., 2013; Rohde et al., 2014; Rubin et al., 2013b). The *nox* target is suitable for *Brachyspira* identification because it is different enough to distinguish between different species, yet is conserved and found ubiquitously among *Brachyspira*, and can be amplified with genus-specific PCR primers (Rohde et al., 2002). The 939 bp PCR product is, however, larger than the optimal fragment size for sequencing on the MiSeq platform, where only approximately 200 base pairs from each end of the target would be sequenced. Long amplicon targets can be sequenced in small

pieces and subsequently assembled into contigs often by shearing the larger amplicon into smaller fragments (Feng et al., 2016; Karlsson and Westerdahl, 2013; Kozich et al., 2013). However, the *nox* amplicon is not sufficiently large for shearing, and so this method may not be suitable. Instead, we investigated each end of the *nox* amplicon sequence to use the ends for identification, rather than relying on the full amplicon sequence.

Bioinformatically, we determined the resolving power of the first 200 base pairs and last 200 base pairs compared to the full *nox* target. Based on the percent identity of pairwise comparisons (Figure 4.1), the similarity between the regions follows the same trend, although the last 200 base pairs may have slightly less differentiation power. Pairwise comparisons of 70 unique *nox* sequences demonstrated that pairwise comparisons of the full target had an average identity match of 90%. Pairwise comparisons of the first 200 bp had an average identity of 88%, and the last 200 bp 93%, thus providing a similar level of differentiation as the full *nox* target.

Another challenge of implementing deep sequencing of the *nox* gene is understanding how much of the actual *Brachyspira* community we are detecting. To investigate this issue, synthetic bacterial communities containing known amounts of genomic DNA from species of interest were created and sequenced. The synthetic communities contained a background of fecal DNA with spiked in *Brachyspira* DNA from different species, and each species was spiked in at high, medium or low levels. Sequence based *Brachyspira* profiles from the synthetic communities generally represented what went into the sequencer based on the qPCR results. The communities that had one species spiked in at a high level were dominated by that species. Synthetic communities S9, S10, and S11 had all four *Brachyspira* species spiked in at the same level. These communities illustrated that “*B. hampsonii*” strain 30599 was detected at a lower level than the other species.

This suggests that the *nox* primers, although they are “universal”, may have a negative bias towards “*B. hampsonii*” strain 30599, or clade I. In a study in the USA, “*B. hampsonii*” clade I was isolated more frequently from clinical cases than clade II. In composing a collection of isolates selected to represent “*B. hampsonii*” strains in circulation, the authors included 45 clade I and 21 clade II isolates (Mirajkar et al., 2015). Alternatively, isolates obtained from Canadian clinical cases highlighted a higher prevalence of clade II isolates (Perez et al., 2016). The higher prevalence of Clade II in Canada could be due to a higher level of this Clade in circulation, or the lower detection of Clade I could be due to inefficiencies in detection, especially in mixed infections. When presented with an equal mix of clade I and clade II, clade I could be missed (Figure 4.3). “*B. hampsonii*” strain 30599, or Clade I, may be present more frequently and at higher levels than current diagnostic methods based on *nox* PCR may suggest.

The synthetic communities gave us an understanding of what levels of *Brachyspira* we would be able to detect using this sequencing method. From Mexico and Brazil, samples were dominated mostly by *B. hyodysenteriae*. Some Mexican samples were also dominated by *B. pilosicoli*. As *B. hyodysenteriae* is the historically causative agent of swine dysentery, it is not surprising to have it present at large levels in these communities. Previous reports have identified *Brachyspira* as being present in Brazil. *B. hyodysenteriae* and *B. pilosicoli* were both identified, along with unidentified weakly haemolytic isolates (Barcellos et al., 2000), and mixed infections with multiple *Brachyspira* species (Viott et al., 2013). Currently, there is a lack of information regarding *Brachyspira* prevalence in Mexico and whether any “*B. hampsonii*” strains are present in either Brazil or Mexico. “*B. hampsonii*” has been isolated from pigs in Belgium and Germany (Mahu et al., 2014; Rohde et al., 2014), as well as migrating waterfowl (Aller-Moran et al., 2016;

Martinez-Lobo et al., 2013), including birds in North America (Rubin et al., 2013b). Our finding of “*B. hampsonii*” in swine fecal samples from Mexican pigs is thus perhaps not surprising.

While we were able to detect “*B. hampsonii*” in the clinical cases from Mexico and Brazil through deep sequencing, some were present at extremely low levels. Brazilian cases contained “*B. hampsonii*” at uniformly low levels, ranging from 0.1 to 0.7% of the community, which has the potential to be due to contamination. “*B. hampsonii*” in Mexican cases comprised 0.1% to 29% of the proportion of the communities, with a median of 4%, suggesting that in Mexico, while it may be present it likely is not a driving factor behind the disease or present at clinically significant levels. While “*B. hampsonii*” may not be a dominant cause of disease in Mexico at this time, there is potential for its emergence as a clinically and economically significant pathogen considering its proximity to the United States. A limitation of the samples used in this study is that there were no DNA extractions controls included, and DNA extractions were used as received. It cannot be ruled out that there is the possibility of laboratory contamination during preparation of samples for PCR and sequencing in our laboratory.

Samples from Canadian clinical cases vary greatly in their *Brachyspira* composition. They have more variation compared to Mexico and Brazil, and the Canadian cases are dominated by a range of different species. Differences between the countries could reflect *Brachyspira* species distribution and/or differences in management practices between countries. For example, in Canada pigs are raised in different barns and/or farms during different stages of their growth whereas in Mexico, the pigs included in this study were raised on farrow-to-finish farms.

All clinical cases, from all three countries, contained multiple *Brachyspira* species though one species typically dominated the community. Traditionally diagnostics attribute one pathogenic

species as being responsible for the disease outcome, though there have been reports of multiple *Brachyspira* species being detected from a single sample (Komarek et al., 2009; La et al., 2016b; Rasback et al., 2005; Rohde et al., 2002; Viott et al., 2013). These communities with a variety of *Brachyspira* species present demonstrate it is possible that multiple species could contribute to the disease outcome. During inoculation trials of “*B. hampsonii*”, not all infected pigs displayed clinical symptoms of swine dysentery (Costa et al., 2014b; Rubin et al., 2013a). “*B. hampsonii*” alone may not have full pathogenic power, but being complemented with other *Brachyspira* spp., including traditionally “non-pathogenic” species, may provide support and contribute to disease pathogenesis. Alternatively, the presence of multiple species may reflect a gut microbiome environment suitable for the pathogenic species and for similar selection of its close relatives, which may simply be non-pathogenic bystanders. Multiple species inoculation trials, as opposed to single species trials, would provide insight as to whether multiple species contribute to the pathogenesis of swine dysentery.

4.6 Conclusions

Deep sequencing of the *nox* gene target is a suitable method to describe the composition of *Brachyspira* species in the fecal microbial community. The technique was used to investigate synthetic *Brachyspira* communities and clinical cases of swine mucohaemorrhagic diarrhea from Western Canada, Mexico and Brazil. Pigs with disease were frequently colonized by multiple species of *Brachyspira*. “*Brachyspira hampsonii*” was detected in both Mexican and Brazilian samples though at low levels and communities were not dominated by this species, suggesting it

may not be clinically significant. Synthetic *Brachyspira* communities revealed a negative primer bias towards “*B. hampsonii*” strain 30599 (Clade I), which has implications for diagnostic detection of these pathogens.

5 General discussion

5.1 Summary and limitations of these works

5.1.1 Comparative genomics of “*B. hampsonii*” strains reveals gene content differences

“*B. hampsonii*” is a novel *Brachyspira* species that is associated with the re-emergence of *Brachyspira*-associated disease in Canada and the United States. Genome sequences for pathogenic “*B. hampsonii*” strain 30446 and non-pathogenic “*B. hampsonii*” strain KL-180 were available, and allowed for comparative genomics approach to be used to identify gene content differences, which could correlate to putative virulence factors.

Chapter two highlights gene content differences identified between pathogenic and non-pathogenic *Brachyspira*. The genome annotation through the RAST server identified 25 genes unique to the pathogenic strain 30446 and 11 genes unique to the non-pathogenic strain KL-180. The annotation from JGI-MGAP revealed a set of 25 genes, unique to pathogenic strains of “*B. hampsonii*” 30446 and 30599, as well as *B. hyodysenteriae*, but absent in the “*B. hampsonii*” KL-180. Very similar distributions of COGs were observed between the genomes, suggesting the differing pathogenicity of the strains is due to small genomic differences. However, only 50-59% of the genes from each genome were classified into COG categories. *Brachyspira* is not well characterized, making identifying genes difficult. Many differences between the genomes correspond to “hypothetical proteins”, making it impossible to know whether those differences could be important to *Brachyspira*-associated disease.

One set of genes found to be present in the pathogenic “*B. hampsonii*” strain 30446 and absent in the non-pathogenic “*B. hampsonii*” strain KL-180 through genome annotation corresponds to the *sag* operon. PCR assays were created to screen epidemiologically distinct “*B. hampsonii*” strains and *Brachyspira* ATCC strains for the *sag* genes. *sagA* was not detected in “*B. hampsonii*” strains D12-32529 and D13-06010, and *sagG* and *sagI* were not detected in “*B. hampsonii*” strain D13-06010. All the genes were detected in ATCC strains for *B. murdochii*, *B. innocens* and *B. hyodysenteriae*, and *sagA* and *sagD* were not detected in *B. pilosicoli*. It was unexpected to detect the *sag* genes species considered non-pathogenic, such as *B. innocens*. However, this study only looked for the detection of ORFs and did not explore expression or function of *sag* genes. It is possible the genes are present in non-pathogenic *Brachyspira* strains though may not be expressed.

The comparative genomics utilized a pathogenic and non-pathogenic strain of “*B. hampsonii*”. The non-pathogenic “*B. hampsonii*” strain KL-180 was isolated from a lesser snow goose, and through swine inoculation trials, was demonstrated to be non-pathogenic in pigs. The pathogenicity of the other “*B. hampsonii*” strains that were screened for the *sag* genes (besides strains 30446 and 30599) have not been experimentally demonstrated to be pathogenic. Their pathogenicity is assumed because they were isolated from pigs with disease. The comparative genomics would have been strengthened with more non-pathogenic representatives. Additional isolates, similar to KL-180, were not viable after freezer storage. Additional *Brachyspira* isolates could be obtained from snow geese and other reservoirs (such as rodents), to hopefully isolate additional non-pathogenic “*B. hampsonii*” strains which would strengthen the comparative genomics analysis.

5.1.2 Deep sequencing of fecal DNA from clinical cases of *Brachyspira*-associated disease reveal multiple *Brachyspira* spp. infections are common

The results in Chapter 3 demonstrated that mixed *Brachyspira* spp. infections are common in clinical cases of mucohaemorrhagic diarrhea. Deep sequencing of cases from Western Canada, Brazil, and Mexico showed that *Brachyspira* profiles from each country differed. Brazilian profiles were dominated by *B. hyodysenteriae*. Mexican profiles were dominated by *B. hyodysenteriae* and *B. pilosicoli*, and Canadian samples were dominated by a range of *Brachyspira* spp.. Canadian profiles included samples dominated by “*B. hampsonii*”, the species which has been an important factor in the re-emergence of disease in Canada and the United States. Samples obtained for this study were used as they were received, and did not include DNA extraction controls which is a significant limitation of this study. “*B. hampsonii*” was detected at extremely low levels in the Brazilian samples and some Mexican samples, but it could likely be due to contamination. With DNA extraction controls, contamination during processing in the lab could have been better accounted for.

5.2 Discussion of future prospects

The work described in Chapter 2 focused on identifying putative virulence factors using a comparative genomics approach. This work is relevant as much about *Brachyspira* pathogenesis and virulence factors are not currently understood. Improved understanding of the genomic components of *Brachyspira* would aid in identifying putative virulence factors. Many genes identified through the comparative genomics approach were labelled as hypothetical proteins, and as such it is difficult to identify important components. Additionally, studying the expression of

genes identified as putative virulence factors will aid the understanding of their importance in *Brachyspira* infection. One set of genes that would benefit from such work is the *sag*-like genes identified in pathogenic *Brachyspira* in Chapter 2.

In Chapter 3, deep sequencing of fecal DNA from pigs with mucohaemorrhagic diarrhea showed that mixed *Brachyspira* infections are common. Current diagnostic methods often attribute disease with one pathogenic *Brachyspira* species, though it is possible multiple species could play a role in disease outcome. The pathogenicity of single *Brachyspira* isolates has been demonstrated through inoculation trials for many species. During these inoculation trials, pathogenic *Brachyspira* species do not universally cause disease and a proportion of pigs do not show clinical signs. Inoculation trials with multiple *Brachyspira* species at once would help describe whether multiple species, or specific combinations of species are important to disease development.

Additionally, in this chapter a new method of deep sequencing of the *nox* gene target was described which could be a useful diagnostic tool. This method allows for simultaneous detection of all *Brachyspira* species, rather than identification of a single species. With sequencing costs becoming economical, this may be a useful tool for swine producers, veterinarians and researchers.

6 References

- Aller-Moran, L.M., Martinez-Lobo, F.J., Rubio, P., Carvajal, A., 2016. Experimental infection of conventional pigs with a '*Brachyspira hampsonii*' isolate recovered from a migrating waterfowl in Spain. *The Veterinary Journal* 214, 10-13.
- Allewelt, M., Coleman, F.T., Grout, M., Priebe, G.P., Pier, G.B., 2000. Acquisition of expression of the *Pseudomonas aeruginosa* ExoU cytotoxin leads to increased bacterial virulence in a murine model of acute pneumonia and systemic spread. *Infection and Immunity* 68, 3998-4004.
- Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O., 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9, 75.
- Backhans, A., Johansson, K.E., Fellstrom, C., 2010. Phenotypic and molecular characterization of *Brachyspira* spp. isolated from wild rodents. *Environmental Microbiology Reports* 2, 720-727.
- Barcellos, D.E., Mathiesen, M.R., de Uzeda, M., Kader, II, Duhamel, G.E., 2000. Prevalence of *Brachyspira* species isolated from diarrhoeic pigs in Brazil. *Veterinary Record* 146, 398-403.
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., Horvath, P., 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709-1712.
- Bellgard, M.I., Wanchanthuek, P., La, T., Ryan, K., Moolhuijzen, P., Albertyn, Z., Shaban, B., Motro, Y., Dunn, D.S., Schibeci, D., Hunter, A., Barrero, R., Phillips, N.D., Hampson, D.J., 2009. Genome sequence of the pathogenic intestinal spirochete *Brachyspira hyodysenteriae* reveals adaptations to its lifestyle in the porcine large intestine. *PLoS One* 4, e4641.
- Bizzini, A., Greub, G., 2010. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry, a revolution in clinical microbial identification. *Clinical Microbiology and Infection* 16, 1614-1619.
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114-2120.

- Boye, M., Baloda, S.B., Leser, T.D., Moller, K., 2001. Survival of *Brachyspira hyodysenteriae* and *B. pilosicoli* in terrestrial microcosms. *Veterinary microbiology* 81, 33-40.
- Burrough, E., Strait, E., Kinyon, J., Bower, L., Madson, D., Schwartz, K., Frana, T., Songer, J.G., 2012a. Comparison of atypical *Brachyspira* spp. clinical isolates and classic strains in a mouse model of swine dysentery. *Veterinary microbiology* 160, 387-394.
- Burrough, E.R., 2016. Swine Dysentery: Etiopathogenesis and Diagnosis of a Reemerging Disease. *Veterinary Pathology*.
- Burrough, E.R., Strait, E.L., Kinyon, J.M., Bower, L.P., Madson, D.M., Wilberts, B.L., Schwartz, K.J., Frana, T.S., Songer, J.G., 2012b. Comparative virulence of clinical *Brachyspira* spp. isolates in inoculated pigs. *Journal of Veterinary Diagnostic Investigation* 24, 1025-1034.
- Calderaro, A., Piccolo, G., Montecchini, S., Buttrini, M., Gorrini, C., Rossi, S., Arcangeletti, M.C., De Conto, F., Medici, M.C., Chezzi, C., 2013. MALDI-TOF MS analysis of human and animal *Brachyspira* species and benefits of database extension. *Journal of Proteomics* 78, 273-280.
- Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunencko, T., Zaneveld, J., Knight, R., 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* 7, 335-336.
- Caspi, R., Foerster, H., Fulcher, C.A., Kaipa, P., Krummenacker, M., Latendresse, M., Paley, S., Rhee, S.Y., Shearer, A.G., Tissier, C., Walk, T.C., Zhang, P., Karp, P.D., 2008. The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Research* 36, D623-631.
- Censini, S., Lange, C., Xiang, Z., Crabtree, J.E., Ghiara, P., Borodovsky, M., Rappuoli, R., Covacci, A., 1996. cag, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. *Proceedings of the National Academy of Sciences of the United States of America* 93, 14648-14653.
- Chander, Y., Primus, A., Oliveira, S., Gebhart, C.J., 2012. Phenotypic and molecular characterization of a novel strongly hemolytic *Brachyspira* species, provisionally designated "*Brachyspira hampsonii*". *Journal of Veterinary Diagnostic Investigation* 24, 903-910.
- Chia, S.P., Taylor, D.J., 1978. Factors affecting the survival of *Treponema hyodysenteriae* in dysenteric pig faeces. *The Veterinary record* 103, 68-70.

- Costa, M.O., Chaban, B., Harding, J.C., Hill, J.E., 2014a. Characterization of the fecal microbiota of pigs before and after inoculation with "*Brachyspira hampsonii*". PLoS One 9, e106399.
- Costa, M.O., Hill, J.E., Fernando, C., Lemieux, H.D., Detmer, S.E., Rubin, J.E., Harding, J.C., 2014b. Confirmation that "*Brachyspira hampsonii*" clade I (Canadian strain 30599) causes mucohemorrhagic diarrhea and colitis in experimentally infected pigs. BMC Veterinary Research 10, 129.
- Datta, V., Myskowski, S.M., Kwinn, L.A., Chiem, D.N., Varki, N., Kansal, R.G., Kotb, M., Nizet, V., 2005. Mutational analysis of the group A streptococcal operon encoding streptolysin S and its virulence role in invasive infection. Molecular Microbiology 56, 681-695.
- Delcher, A.L., Harmon, D., Kasif, S., White, O., Salzberg, S.L., 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Research 27, 4636-4641.
- Eitinger, T., Mandrand-Berthelot, M.A., 2000. Nickel transport systems in microorganisms. Archives of Microbiology 173, 1-9.
- Erickson, D.L., Jarrett, C.O., Callison, J.A., Fischer, E.R., Hinnebusch, B.J., 2008. Loss of a biofilm-inhibiting glycosyl hydrolase during the emergence of *Yersinia pestis*. Journal of Bacteriology 190, 8163-8170.
- Fellstrom, C., Rasback, T., Johansson, K.E., Olofsson, T., Aspan, A., 2008. Identification and genetic fingerprinting of *Brachyspira* species. Journal of Microbiological Methods 72, 133-140.
- Feng, Y.J., Liu, Q.F., Chen, M.Y., Liang, D., Zhang, P., 2016. Parallel tagged amplicon sequencing of relatively long PCR products using the Illumina HiSeq platform and transcriptome assembly. Molecular Ecology Resources 16, 91-102.
- Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L., Tate, J., Punta, M., 2014. Pfam: the protein families database. Nucleic Acids Research 42, D222-230.
- Gal-Mor, O., Finlay, B.B., 2006. Pathogenicity islands: a molecular toolbox for bacterial virulence. Cellular Microbiology 8, 1707-1719.
- Galperin, M.Y., Makarova, K.S., Wolf, Y.I., Koonin, E.V., 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Research 43, D261-269.
- Georgiades, K., 2012. Genomics of epidemic pathogens. Clinical Microbiology and Infection 18, 213-217.

- Glaser, P., Frangeul, L., Buchrieser, C., Rusniok, C., Amend, A., Baquero, F., Berche, P., Bloecker, H., Brandt, P., Chakraborty, T., Charbit, A., Chetouani, F., Couve, E., de Daruvar, A., Dehoux, P., Domann, E., Dominguez-Bernal, G., Duchaud, E., Durant, L., Dussurget, O., Entian, K.D., Fsihi, H., Garcia-del Portillo, F., Garrido, P., Gautier, L., Goebel, W., Gomez-Lopez, N., Hain, T., Hauf, J., Jackson, D., Jones, L.M., Kaerst, U., Kreft, J., Kuhn, M., Kunst, F., Kurapkat, G., Madueno, E., Maitournam, A., Vicente, J.M., Ng, E., Nedjari, H., Nordsiek, G., Novella, S., de Pablos, B., Perez-Diaz, J.C., Purcell, R., Rammel, B., Rose, M., Schlueter, T., Simoes, N., Tierrez, A., Vazquez-Boland, J.A., Voss, H., Wehland, J., Cossart, P., 2001. Comparative genomics of *Listeria* species. *Science* 294, 849-852.
- Glock, R.D., Harris, D.L., Kluge, J.P., 1974. Localization of spirochetes with the structural characteristics of *Treponema hyodysenteriae* in the lesions of swine dysentery. *Infection and Immunity* 9, 167-178.
- Haft, D.H., Selengut, J.D., White, O., 2003. The TIGRFAMs database of protein families. *Nucleic Acids Research* 31, 371-373.
- Harding, C., Fernando, C., Rubin, J.E., Hill, J.E. 2012. Swine dysentery and other emerging *Brachyspiras*. . In: *Advances in Pork Production*, Banff, Alberta, 127-132.
- Harding, J.C., Hill, J.E., Chirino-Trejo, M., O'Connor, B., Rubin, J.E., Fernando, C. 2010. Detection of a novel *Brachyspira* species associated with haemorrhagic and necrotizing colitis. . In: *International Pig Veterinary Society Congress*, Vancouver, Canada, July 18-21, 2010, 740.
- Harris, D.L., Alexander, T.J., Whipp, S.C., Robinson, I.M., Glock, R.D., Matthews, P.J., 1978. Swine dysentery: studies of gnotobiotic pigs inoculated with *Treponema hyodysenteriae*, *Bacteroides vulgatus*, and *Fusobacterium necrophorum*. *Journal of the Veterinary Medical Association* 172, 468-471.
- Harris, D.L., Glock, R.D., 1971. Swine Dysentery: A Review. *Iowa State University Veterinarian* 33.
- Harris, D.L., Glock, R.D., Christensen, C.R., Kinyon, J.M., 1972. Inoculation of pigs with *Treponema hyodysenteriae* (new species) and reproduction of the disease. *Veterinary Medicine, Small Animal Clinician* 67, 61-64.
- Hemm, M.R., Paul, B.J., Schneider, T.D., Storz, G., Rudd, K.E., 2008. Small membrane proteins found by comparative genomics and ribosome binding site models. *Molecular Microbiology* 70, 1487-1501.
- Hidalgo, A., Rubio, P., Osorio, J., Carvajal, A., 2010. Prevalence of *Brachyspira pilosicoli* and "*Brachyspira canis*" in dogs and their association with diarrhoea. *Veterinary microbiology* 146, 356-360.

- Hsu, T., Hutto, D.L., Minion, F.C., Zuerner, R.L., Wannemuehler, M.J., 2001. Cloning of a beta-hemolysin gene of *Brachyspira (Serpulina) hyodysenteriae* and its expression in *Escherichia coli*. *Infection and Immunity* 69, 706-711.
- Huntemann, M., Ivanova, N.N., Mavromatis, K., Tripp, H.J., Paez-Espino, D., Palaniappan, K., Szeto, E., Pillay, M., Chen, I.M., Pati, A., Nielsen, T., Markowitz, V.M., Kyrpides, N.C., 2015. The standard operating procedure of the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v.4). *Standards in Genomic Sciences* 10, 86.
- Hyatt, D., Chen, G.L., Locascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J., 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11, 119.
- Hyatt, D.R., ter Huurne, A.A., van der Zeijst, B.A., Joens, L.A., 1994. Reduced virulence of *Serpulina hyodysenteriae* hemolysin-negative mutants in pigs and their potential to protect pigs against challenge with a virulent strain. *Infection and Immunity* 62, 2244-2248.
- Jansson, D.S., Fellstrom, C., Rasback, T., Vagsholm, I., Gunnarsson, A., Ingermaa, F., Johansson, K.E., 2008. Phenotypic and molecular characterization of *Brachyspira* spp. isolated from laying hens in different housing systems. *Veterinary microbiology* 130, 348-362.
- Jansson, D.S., Johansson, K.E., Olofsson, T., Rasback, T., Vagsholm, I., Pettersson, B., Gunnarsson, A., Fellstrom, C., 2004. *Brachyspira hyodysenteriae* and other strongly beta-haemolytic and indole-positive spirochaetes isolated from mallards (*Anas platyrhynchos*). *Journal of Medical Microbiology* 53, 293-300.
- Jansson, D.S., Mushtaq, M., Johansson, K.E., Bonnedahl, J., Waldenstrom, J., Andersson, D.I., Broman, T., Berg, C., Olsen, B., 2015. Intestinal spirochaetes (genus *Brachyspira*) colonise wild birds in the southern Atlantic region and Antarctica. *Infection Ecology and Epidemiology* 5, 29296.
- Jensen, T.K., Christensen, A.S., Boye, M., 2010. *Brachyspira murdochii* colitis in pigs. *Veterinary Pathology* 47, 334-338.
- Joens, L.A., Kinyon, J.M., 1982. Isolation of *Treponema hyodysenteriae* from wild rodents. *Journal of Clinical Microbiology* 15, 994-997.
- Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., Tanabe, M., 2014. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Research* 42, D199-205.
- Karlsson, M., Westerdahl, H., 2013. Characteristics of MHC class I genes in house sparrows *Passer domesticus* as revealed by long cDNA transcripts and amplicon sequencing. *Journal of Molecular Evolution* 77, 8-21.

- Kennedy, M.J., Yancey, R.J., Jr., 1996. Motility and chemotaxis in *Serpulina hyodysenteriae*. *Veterinary microbiology* 49, 21-30.
- Kinyon, J.M., Harris, D.L., 1979. *Treponema innocens*, a new species of intestinal bacteria, and emended description of the type strain of *Treponema hyodysenteriae*. *International Journal of Systematic Bacteriology* 29, 102-109.
- Komarek, V., Maderner, A., Spargser, J., Weissenböck, H., 2009. Infections with weakly haemolytic *Brachyspira* species in pigs with miscellaneous chronic diseases. *Veterinary microbiology* 134, 311-317.
- Koressaar, T., Remm, M., 2007. Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23, 1289-1291.
- Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., Schloss, P.D., 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and Environmental Microbiology* 79, 5112-5120.
- La, T., Collins, A.M., Phillips, N.D., Oksa, A., Hampson, D.J., 2006. Development of a multiplex-PCR for rapid detection of the enteric pathogens *Lawsonia intracellularis*, *Brachyspira hyodysenteriae*, and *Brachyspira pilosicoli* in porcine faeces. *Letters in Applied Microbiology* 42, 284-288.
- La, T., Phillips, N.D., Hampson, D.J., 2003. Development of a duplex PCR assay for detection of *Brachyspira hyodysenteriae* and *Brachyspira pilosicoli* in pig feces. *Journal of Clinical Microbiology* 41, 3372-3375.
- La, T., Phillips, N.D., Hampson, D.J., 2016a. An investigation into the etiological agents of swine dysentery in Australian pig herds. *PLoS One* 11, e0167424.
- La, T., Rohde, J., Phillips, N.D., Hampson, D.J., 2016b. Comparison of *Brachyspira hyodysenteriae* isolates recovered from pigs in apparently healthy multiplier herds with isolates from herds with swine dysentery. *PLoS One* 11, e0160362.
- Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9, 357-359.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947-2948.
- Lee, S.W., Mitchell, D.A., Markley, A.L., Hensler, M.E., Gonzalez, D., Wohlrab, A., Dorrestein, P.C., Nizet, V., Dixon, J.E., 2008. Discovery of a widely distributed toxin biosynthetic

- gene cluster. Proceedings of the National Academy of Sciences of the United States of America 105, 5879-5884.
- Lery, L.M., Frangeul, L., Tomas, A., Passet, V., Almeida, A.S., Bialek-Davenet, S., Barbe, V., Bengoechea, J.A., Sansonetti, P., Brisse, S., Tournebize, R., 2014. Comparative analysis of *Klebsiella pneumoniae* genomes identifies a phospholipase D family protein as a novel virulence factor. BMC Biology 12, 41.
- Letzel, A.C., Pidot, S.J., Hertweck, C., 2014. Genome mining for ribosomally synthesized and post-translationally modified peptides (RiPPs) in anaerobic bacteria. BMC Genomics 15, 983.
- Loman, N.J., Constantinidou, C., Chan, J.Z., Halachev, M., Sergeant, M., Penn, C.W., Robinson, E.R., Pallen, M.J., 2012. High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. Nature Reviews Microbiology 10, 599-606.
- Lopez-Fernandez, S., Sonogo, P., Moretto, M., Pancher, M., Engelen, K., Pertot, I., Campisano, A., 2015. Whole-genome comparative analysis of virulence genes unveils similarities and differences between endophytes and other symbiotic bacteria. Frontiers in Microbiology 6, 419.
- Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Research 25, 955-964.
- Mahu, M., de Jong, E., De Pauw, N., Vande Maele, L., Vandenbroucke, V., Vandersmissen, T., Miry, C., Pasmans, F., Haesebrouck, F., Martel, A., Boyen, F., 2014. First isolation of "*Brachyspira hampsonii*" from pigs in Europe. Veterinary Record 174, 47.
- Mahu, M., De Pauw, N., Vande Maele, L., Verlinden, M., Boyen, F., Ducatelle, R., Haesebrouck, F., Martel, A., Pasmans, F., 2016. Variation in hemolytic activity of *Brachyspira hyodysenteriae* strains from pigs. Veterinary Research 47, 66.
- Margawani, K.R., Robertson, I.D., Brooke, C.J., Hampson, D.J., 2004. Prevalence, risk factors and molecular epidemiology of *Brachyspira pilosicoli* in humans on the island of Bali, Indonesia. Journal of Medical Microbiology 53, 325-332.
- Martin-Platero, A.M., Valdivia, E., Maqueda, M., Martinez-Bueno, M., 2007. Fast, convenient, and economical method for isolating genomic DNA from lactic acid bacteria using a modification of the protein "salting-out" procedure. Analytical Biochemistry 366, 102-104.
- Martinez-Lobo, F.J., Hidalgo, A., Garcia, M., Arguello, H., Naharro, G., Carvajal, A., Rubio, P., 2013. First identification of "*Brachyspira hampsonii*" in wild European waterfowl. PLoS One 8, e82626.

- Maurelli, A.T., 2007. Black holes, antivirulence genes, and gene inactivation in the evolution of bacterial pathogens. *FEMS Microbiology Letters* 267, 1-8.
- McLaren, A.J., Trott, D.J., Swayne, D.E., Oxberry, S.L., Hampson, D.J., 1997. Genetic and phenotypic characterization of intestinal spirochetes colonizing chickens and allocation of known pathogenic isolates to three distinct genetic groups. *Journal of Clinical Microbiology* 35, 412-417.
- Milner, J.A., Sellwood, R., 1994. Chemotactic response to mucin by *Serpulina hyodysenteriae* and other porcine spirochetes: potential role in intestinal colonization. *Infection and Immunity* 62, 4095-4099.
- Mirajkar, N.S., Bekele, A.Z., Chander, Y.Y., Gebhart, C.J., 2015. Molecular epidemiology of novel pathogen "*Brachyspira hampsonii*" reveals relationships between diverse genetic groups, regions, host species, and other pathogenic and commensal *Brachyspira* species. *Journal of Clinical Microbiology* 53, 2908-2918.
- Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., Soria, E., 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of Molecular Evolution* 60, 174-182.
- Molloy, E.M., Cotter, P.D., Hill, C., Mitchell, D.A., Ross, R.P., 2011. Streptolysin S-like virulence factors: the continuing *sagA*. *Nature Reviews Microbiology* 9, 670-681.
- Muir, S., Koopman, M.B., Libby, S.J., Joens, L.A., Heffron, F., Kusters, J.G., 1992. Cloning and expression of a *Serpula* (*Treponema*) *hyodysenteriae* hemolysin gene. *Infection and Immunity* 60, 529-535.
- Naresh, R., Hampson, D.J., 2010. Attraction of *Brachyspira pilosicoli* to mucin. *Microbiology* 156, 191-197.
- Nibbelink, S.K., Wannemuehler, M.J., 1991. Susceptibility of inbred mouse strains to infection with *Serpula* (*Treponema*) *hyodysenteriae*. *Infection and Immunity* 59, 3111-3118.
- Nishibuchi, M., Kaper, J.B., 1995. Thermostable direct hemolysin gene of *Vibrio parahaemolyticus*: a virulence gene acquired by a marine bacterium. *Infection and Immunity* 63, 2093-2099.
- Nizet, V., Beall, B., Bast, D.J., Datta, V., Kilburn, L., Low, D.E., De Azavedo, J.C., 2000. Genetic locus for streptolysin S production by group A streptococcus. *Infection and Immunity* 68, 4245-4254.
- Nuessen, M.E., Birmingham, J.R., Joens, L.A., 1982. Biological activity of a lipopolysaccharide extracted from *Treponema hyodysenteriae*. *Infection and Immunity* 37, 138-142.

- Nuessen, M.E., Joens, L.A., Glock, R.D., 1983. Involvement of lipopolysaccharide in the pathogenicity of *Treponema hyodysenteriae*. *Journal of Immunology* 131, 997-999.
- Ochiai, S., Adachi, Y., Mori, K., 1997. Unification of the genera *Serpulina* and *Brachyspira*, and proposals of *Brachyspira hyodysenteriae* Comb. Nov., *Brachyspira innocens* Comb. Nov. and *Brachyspira pilosicoli* Comb. Nov. *Microbiology and Immunology* 41, 445-452.
- Osorio, J., Carvajal, A., Naharro, G., Rubio, P., La, T., Phillips, N.D., Hampson, D.J., 2013. Identification of weakly haemolytic *Brachyspira* isolates recovered from pigs with diarrhoea in Spain and Portugal and comparison with results from other countries. *Research in Veterinary Science* 95, 861-869.
- Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.Y., Cohoon, M., de Crecy-Lagard, V., Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E.D., Gerdes, S., Glass, E.M., Goesmann, A., Hanson, A., Iwata-Reuyl, D., Jensen, R., Jamshidi, N., Krause, L., Kubal, M., Larsen, N., Linke, B., McHardy, A.C., Meyer, F., Neuweyer, H., Olsen, G., Olson, R., Osterman, A., Portnoy, V., Pusch, G.D., Rodionov, D.A., Ruckert, C., Steiner, J., Stevens, R., Thiele, I., Vassieva, O., Ye, Y., Zagnitko, O., Vonstein, V., 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Research* 33, 5691-5702.
- Oxberry, S.L., Trott, D.J., Hampson, D.J., 1998. *Serpulina pilosicoli*, waterbirds and water: potential sources of infection for humans and other animals. *Epidemiology and Infection* 121, 219-225.
- Patterson, A.H., Rubin, J.E., Fernando, C., Costa, M.O., Harding, J.C., Hill, J.E., 2013. Fecal shedding of *Brachyspira* spp. on a farrow-to-finish swine farm with a clinical history of "*Brachyspira hampsonii*"-associated colitis. *BMC Veterinary Research* 9, 137.
- Perez, J.B.D.S., Rubin, J.E., Fernando, C., Harding, J.C., Hill, J.E., 2016. Characterization of "*Brachyspira hampsonii*" Clade I and II isolated from commercial swine in Western Canada. *FACETS* 1, 163-172.
- Petkau, A., Stuart-Edwards, M., Stothard, P., Van Domselaar, G., 2010. Interactive microbial genome visualization with GView. *Bioinformatics* 26, 3125-3126.
- Phillips, N.D., La, T., Adams, P.J., Harland, B.L., Fenwick, S.G., Hampson, D.J., 2009. Detection of *Brachyspira hyodysenteriae*, *Lawsonia intracellularis* and *Brachyspira pilosicoli* in feral pigs. *Veterinary microbiology* 134, 294-299.
- Phillips, N.D., La, T., Amin, M.M., Hampson, D.J., 2010. *Brachyspira intermedia* strain diversity and relationships to the other indole-positive *Brachyspira* species. *Veterinary microbiology* 143, 246-254.

- Rasback, T., Fellstrom, C., Bergsjö, B., Cizek, A., Collin, K., Gunnarsson, A., Jensen, S.M., Mars, A., Thomson, J., Vyt, P., Pringle, M., 2005. Assessment of diagnostics and antimicrobial susceptibility testing of *Brachyspira* species using a ring test. *Veterinary microbiology* 109, 229-243.
- Rasback, T., Jansson, D.S., Johansson, K.E., Fellstrom, C., 2007. A novel enteropathogenic, strongly haemolytic spirochaete isolated from pig and mallard, provisionally designated '*Brachyspira suanatina*' sp. nov. *Environmental Microbiology* 9, 983-991.
- Richter, M., Rossello-Mora, R., 2009. Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences of the United States of America* 106, 19126-19131.
- Rohde, J., Habighorst-Blome, K., Seehusen, F., 2014. "*Brachyspira hampsonii*" clade I isolated from Belgian pigs imported to Germany. *Veterinary microbiology* 168, 432-435.
- Rohde, J., Rothkamp, A., Gerlach, G.F., 2002. Differentiation of porcine *Brachyspira* species by a novel *nox* PCR-based restriction fragment length polymorphism analysis. *Journal of Clinical Microbiology* 40, 2598-2600.
- Rosey, E.L., Kennedy, M.J., Yancey, R.J., Jr., 1996. Dual *flaA1 flaB1* mutant of *Serpulina hyodysenteriae* expressing periplasmic flagella is severely attenuated in a murine model of swine dysentery. *Infection and Immunity* 64, 4154-4162.
- Rubin, J.E., Costa, M.O., Hill, J.E., Kittrell, H.E., Fernando, C., Huang, Y., O'Connor, B., Harding, J.C., 2013a. Reproduction of mucohaemorrhagic diarrhea and colitis indistinguishable from swine dysentery following experimental inoculation with "*Brachyspira hampsonii*" strain 30446. *PLoS One* 8, e57146.
- Rubin, J.E., Harms, N.J., Fernando, C., Soos, C., Detmer, S.E., Harding, J.C., Hill, J.E., 2013b. Isolation and characterization of *Brachyspira* spp. including "*Brachyspira hampsonii*" from lesser snow geese (*Chen caerulescens caerulescens*) in the Canadian Arctic. *Microbial Ecology* 66, 813-822.
- Song, Y., Hampson, D.J., 2009. Development of a multiplex qPCR for detection and quantitation of pathogenic intestinal spirochaetes in the faeces of pigs and chickens. *Veterinary microbiology* 137, 129-136.
- Stanton, T.B., 1992. Proposal to change the genus designation *Serpula* to *Serpulina* gen. nov. containing the species *Serpulina hyodysenteriae* comb. nov. and *Serpulina innocens* comb. nov. *International Journal of Systematic Bacteriology* 42, 189-190.
- Stanton, T.B., Jensen, N.S., Casey, T.A., Tordoff, L.A., Dewhirst, F.E., Paster, B.J., 1991. Reclassification of *Treponema hyodysenteriae* and *Treponema innocens* in a new genus,

- Serpula* gen. nov., as *Serpula hyodysenteriae* comb. nov. and *Serpula innocens* comb. nov. International Journal of Systematic Bacteriology 41, 50-58.
- Stanton, T.B., Rosey, E.L., Kennedy, M.J., Jensen, N.S., Bosworth, B.T., 1999. Isolation, oxygen sensitivity, and virulence of NADH oxidase mutants of the anaerobic spirochete *Brachyspira* (*Serpulina*) *hyodysenteriae*, etiologic agent of swine dysentery. Applied and Environmental Microbiology 65, 5028-5034.
- Stein, L., 2001. Genome annotation: from sequence to biology. Nature Reviews Genetics 2, 493-503.
- Stephens, C.P., Hampson, D.J., 2002. Experimental infection of broiler breeder hens with the intestinal spirochaete *Brachyspira* (*Serpulina*) *pilosicoli* causes reduced egg production. Avian Pathology 31, 169-175.
- Tatusov, R.L., Natale, D.A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S., Kiryutin, B., Galperin, M.Y., Fedorova, N.D., Koonin, E.V., 2001. The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Research 29, 22-28.
- ter Huurne, A.A., Muir, S., van Houten, M., van der Zeijst, B.A., Gaastra, W., Kusters, J.G., 1994. Characterization of three putative *Serpulina hyodysenteriae* hemolysins. Microbial Pathology 16, 269-282.
- Trott, D.J., Huxtable, C.R., Hampson, D.J., 1996. Experimental infection of newly weaned pigs with human and porcine strains of *Serpulina pilosicoli*. Infection and Immunity 64, 4648-4654.
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., Rozen, S.G., 2012. Primer3--new capabilities and interfaces. Nucleic Acids Research 40, e115.
- Viott, A.M., Lage, A.P., Cruz, E.C., Jr., Guedes, R.M., 2013. The prevalence of swine enteropathogens in Brazilian grower and finish herds. Brazilian Journal of Microbiology 44, 145-151.
- Warneke, H.L., Kinyon, J.M., Bower, L.P., Burrough, E.R., Frana, T.S., 2014. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry for rapid identification of *Brachyspira* species isolated from swine, including the newly described "*Brachyspira hampsonii*". Journal of Veterinary Diagnostic Investigation 26, 635-639.
- Warren, A.S., Archuleta, J., Feng, W.C., Setubal, J.C., 2010. Missing genes in the annotation of prokaryotic genomes. BMC Bioinformatics 11, 131.
- Weissenböck, H., Maderner, A., Herzog, A.M., Lussy, H., Nowotny, N., 2005. Amplification and sequencing of *Brachyspira* spp. specific portions of nox using paraffin-embedded tissue

- samples from clinical colitis in Austrian pigs shows frequent solitary presence of *Brachyspira murdochii*. *Veterinary microbiology* 111, 67-75.
- Whipp, S.C., Robinson, I.M., Harris, D.L., Glock, R.D., Matthews, P.J., Alexander, T.J., 1979. Pathogenic synergism between *Treponema hyodysenteriae* and other selected anaerobes in gnotobiotic pigs. *Infection and Immunity* 26, 1042-1047.
- Whiting, R.A., Doyle, L.P., Spray, R.S., 1921. Swine Dysentery. *Bulletin*, 3-15.
- Wilberts, B.L., Arruda, P.H., Kinyon, J.M., Madson, D.M., Frana, T.S., Burrough, E.R., 2014. Comparison of session severity, distribution, and colonic mucin expression in pigs with acute swine dysentery following oral inoculation with "*Brachyspira hampsonii*" or *Brachyspira hyodysenteriae*. *Veterinary Pathology* 51, 1096-1108.
- Zhao, S., Guo, Y., Sheng, Q., Shyr, Y., 2014. Advanced heat map and clustering analysis using heatmap3. *BioMed Research International* 2014, 986048.